



```
Db 1 GAAGTCAGAGGCGCTTCAGATTTGTTGGAGACTCGAAAAAAGAGGGGAGCCGAGGC 60
QY 61 AGACGACGGGGGCGGATGCTGCTGTTCTTGGGAGGCGTCTAGCTTTGGCAGCGCGGC 120
Db 61 AGACGACGGGGGCGGATGCTGCTGTTCTTGGGAGGCGTCTAGCTTTGGCAGCGCGGC 120
QY 121 CGCTTTTCTCTCTGGGTGGGCGCGGAGCTCCCGAGTTTGGAGCGCAATTTTTTACAT 180
Db 121 CGCTTTTCTCTCTGGGTGGGCGCGGAGCTCCCGAGTTTGGAGCGCAATTTTTTACAT 180
QY 181 TTTATGGCGATGGGTCAGGCGTTTATCTAGGCGCTCTGGAGGGTACATTTGAAGATGTG 240
Db 181 TTTATGGCGATGGGTCAGGCGTTTATCTAGGCGCTCTGGAGGGTACATTTGAAGATGTG 240
QY 241 CCACCACTCCAAACCGACCAACCTCTGATCTGAGCATGCCCTCATGCTCTCCTTCATGCC 300
Db 241 CCACCACTCCAAACCGACCAACCTCTGATCTGAGCATGCCCTCATGCTCTCCTTCATGCC 300
QY 301 TCCCTTTGGGTGAGGTCAATGTCCTTGGGCGGAGTGGCTTCCCGTTTGAAGCAAGTAT 360
Db 301 TCCCTTTGGGTGAGGTCAATGTCCTTGGGCGGAGTGGCTTCCCGTTTGAAGCAAGTAT 360
QY 361 AATAAGTCTTAGTCAGCTGGCTATAGATGTTCCACATCAGCAAAATCCTTAAACTGGAGG 420
Db 361 AATAAGTCTTAGTCAGCTGGCTATAGATGTTCCACATCAGCAAAATCCTTAAACTGGAGG 420
QY 421 AGAAGAAAGTAGAGTAGAGGGGCGTTCGGCGCTTCGTCAATCGCTAGCGATAGCACAA 480
Db 421 AGAAGAAAGTAGAGTAGAGGGGCGTTCGGCGCTTCGTCAATCGCTAGCGATAGCACAA 480
QY 481 GCTCCCATCGAATCGAGCGCAACATCGACCCGACAAATGACTAAAGCGCAAAACGCCAGCCA 540
Db 481 GCTCCCATCGAATCGAGCGCAACATCGACCCGACCAATGACTAAAGCGCAAAACGCCAGCCA 540
QY 541 ATCAGTATGCCCTTCTCTGCAATCTTTCTCATGCAAGCATTTAAATCTATAGCTAACTA 600
Db 541 ATCAGTATGCCCTTCTCTGCAATCTTTCTCATGCAAGCATTTAAATCTATAGCTAACTA 600
QY 601 CAGCGAGTTTATTAATAACAGGCTATATAGCTGACCTGGCAGTGCTATAGAGCGGCA 660
Db 601 CAGCGAGTTTATTAATAACAGGCTATATAGCTGACCTGGCAGTGCTATAGAGCGGCA 660
QY 661 GCCGCTCTCTTCTATAGCTTTGCTTATGCTATACATCTGTGTGAGCAGTCGATGATTC 720
Db 661 GCCGCTCTCTTCTATAGCTTTGCTTATGCTATACATCTGTGTGAGCAGTCGATGATTC 720
QY 721 AAACAACAATCCGGGCGTTCAGCAAGTCGGAATGAAATTTCCGGCTCATCTCATTTGTCG 780
Db 721 AAACAACAATCCGGGCGTTTCCAGCAAGTCGGAATGAAATTTCCGGCTCATCTCATTTGTCG 780
QY 781 TGGGCTTCACGCGTATTCGCTTAACCGTGTGTAATCAGACCTTCAGAAAGCGCACGCTC 840
Db 781 TGGGCTTCACGCGTATTCGCTTAACCGTGTGTAATCAGACCTTCAGAAAGCGCACGCTC 840
QY 841 CAGCGACCGGTTACCAAGCTTAAAGGCAAAAGGCAAAAGGCAAAAGGCAAAAGGCAAAAG 900
Db 841 CAGCGACCGGTTACCAAGCTTAAAGGCAAAAGGCAAAAGGCAAAAGGCAAAAGGCAAAAG 900
QY 901 CATCTGAACCGTTCAACAGCGCCACGTAATTTCCGGCACGACGCAAGGGCATATCCGTCA 960
Db 901 CATCTGAACCGTTCAACAGCGCCACGTAATTTCCGGCACGACGCAAGGGCATATCCGTCA 960
QY 961 TAGCGAGCGCATAAATTTCTGATTCCTGCTCCGCGCAATTTATCTTTGGGAGGC 1020
Db 961 TAGCGAGCGCATAAATTTCTGATTCCTGCTCCGCGCAATTTATCTTTGGGAGGC 1020
QY 1021 GGGCGGGAATGGAGACAGAGCCCAAGGCAACAAAGTGGCGGTGAGAAATCAACA 1080
Db 1021 GGGCGGGAATGGAGACAGAGCCCAAGGCAACAAAGTGGCGGTGAGAAATCAACA 1080
QY 1081 AGCGGTCTTCCGAGAGAGAGAGAGAG 1111
```

```
Db 1081 AGCGGTCTTCCGAGAGAGAGAGAGAGAG 1111
RESULT 2
AX047682 1111 bp DNA linear PAT 15-DEC-2000
LOCUS
DEFINITION
Sequence 2 from Patent WO0070065.
AX047682
ACCESSION
AX047682.1 GI:11876717
VERSION
KEYWORDS
SOURCE
ORGANISM
Triticum aestivum (bread wheat)
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Triticum.
1
REFERENCE
Gautier,M.F., Thorai,T. and Joudrier,P.
AUTHORS
Promoter of thiredoxine tatrxd2 in wheat
TITLE
Patent: WO 0070065-A 2 23-NOV-2000;
JOURNAL
INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (INRA) (FR)
FEATURES
Location/Qualifiers
source
1..1111
/organism="Triticum aestivum"
/mol_type="unassigned DNA"
/db_xref="taxon:4565"
ORIGIN
Query Match 100.0%; Score 1111; DB 6; Length 1111;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAAGTCAGAGGCGCTTCAGAAATTTGTTGGAGGACTCGAAAAAAGAGGGGAGCCGAGGC 60
Db 1 GAAGTCAGAGGCGCTTCAGAAATTTGTTGGAGGACTCGAAAAAAGAGGGGAGCCGAGGC 60
QY 61 AGACGACGGGGGCGGATGTCCTGTTCTTGGGAGGCGTCTAGCTTTGGCAGCGCGGC 120
Db 61 AGACGACGGGGGCGGATGTCCTGTTCTTGGGAGGCGTCTAGCTTTGGCAGCGCGGC 120
QY 121 CGCTTTTCTCTCTGGGTGGGCGCGGAGCTCCCGAGTTTGGAGCGCAATTTTTTACAT 180
Db 121 CGCTTTTCTCTCTGGGTGGGCGCGGAGCTCCCGAGTTTGGAGCGCAATTTTTTACAT 180
QY 181 TTTATGGCGATGGGTCAGGCGTTTATCTAGGCGCTCTGGAGGGTACATTTGAAGATGTG 240
Db 181 TTTATGGCGATGGGTCAGGCGTTTATCTAGGCGCTCTGGAGGGTACATTTGAAGATGTG 240
QY 241 CCACCAACTCCAAACCGACCAACCTCTGATCTGAGCATGCCCTCATGCTCTCCTTCATGCC 300
Db 241 CCACCAACTCCAAACCGACCAACCTCTGATCTGAGCATGCCCTCATGCTCTCCTTCATGCC 300
QY 301 TCCCTTTGGGTGAGGTCAATGTCCTTGGGCGGAGTGGCTTCCCGTTTGAAGCAAGTAT 360
Db 301 TCCCTTTGGGTGAGGTCAATGTCCTTGGGCGGAGTGGCTTCCCGTTTGAAGCAAGTAT 360
QY 361 AATAAGTCTTAGTCAGCTGGCTATAGATGTTTCCACATCAGCAAAATCCTTAAACTGGAGG 420
Db 361 AATAAGTCTTAGTCAGCTGGCTATAGATGTTTCCACATCAGCAAAATCCTTAAACTGGAGG 420
QY 421 AGAAGAAAGTAGAGTAGAGGGGCGTTCGGCGCTTCGTCAATCGCTAGCGATAGCACAA 480
Db 421 AGAAGAAAGTAGAGTAGAGGGGCGTTCGGCGCTTCGTCAATCGCTAGCGATAGCACAA 480
QY 481 GCTCCCATCGAATCGAGCGCAACATCGACCCGACAAATGACTAAAGCGCAAAACGCCAGCCA 540
Db 481 GCTCCCATCGAATCGAGCGCAACATCGACCCGACCAATGACTAAAGCGCAAAACGCCAGCCA 540
QY 541 ATCAGTATGCCCTTCTCTGCAATCTTTCTCATGCAAGCATTTAAATCTATAGCTAACTA 600
Db 541 ATCAGTATGCCCTTCTCTGCAATCTTTCTCATGCAAGCATTTAAATCTATAGCTAACTA 600
QY 601 CAGCGAGTTTATTAATAACAGGCTATATAGCTGACCTGGCAGTGCTATAGAGCGGCA 660
Db 601 CAGCGAGTTTATTAATAACAGGCTATATAGCTGACCTGGCAGTGCTATAGAGCGGCA 660
QY 661 GCCGCTCTCTTCTATAGCTTTGCTTATGCTATACATCTGTGTGAGCAGTCGATGATTC 720
Db 661 GCCGCTCTCTTCTATAGCTTTGCTTATGCTATACATCTGTGTGAGCAGTCGATGATTC 720
QY 721 AAACAACAATCCGGGCGTTCAGCAAGTCGGAATGAAATTTCCGGCTCATCTCATTTGTCG 780
Db 721 AAACAACAATCCGGGCGTTTCCAGCAAGTCGGAATGAAATTTCCGGCTCATCTCATTTGTCG 780
QY 781 TGGGCTTCACGCGTATTCGCTTAACCGTGTGTAATCAGACCTTCAGAAAGCGCACGCTC 840
Db 781 TGGGCTTCACGCGTATTCGCTTAACCGTGTGTAATCAGACCTTCAGAAAGCGCACGCTC 840
QY 841 CAGCGACCGGTTACCAAGCTTAAAGGCAAAAGGCAAAAGGCAAAAGGCAAAAGGCAAAAG 900
Db 841 CAGCGACCGGTTACCAAGCTTAAAGGCAAAAGGCAAAAGGCAAAAGGCAAAAGGCAAAAG 900
QY 901 CATCTGAACCGTTCAACAGCGCCACGTAATTTCCGGCACGACGCAAGGGCATATCCGTCA 960
Db 901 CATCTGAACCGTTCAACAGCGCCACGTAATTTCCGGCACGACGCAAGGGCATATCCGTCA 960
QY 961 TAGCGAGCGCATAAATTTCTGATTCCTGCTCCGCGCAATTTATCTTTGGGAGGC 1020
Db 961 TAGCGAGCGCATAAATTTCTGATTCCTGCTCCGCGCAATTTATCTTTGGGAGGC 1020
QY 1021 GGGCGGGAATGGAGACAGAGCCCAAGGCAACAAAGTGGCGGTGAGAAATCAACA 1080
Db 1021 GGGCGGGAATGGAGACAGAGCCCAAGGCAACAAAGTGGCGGTGAGAAATCAACA 1080
QY 1081 AGCGGTCTTCCGAGAGAGAGAGAGAGAG 1111
```

Db	601	CACCCAGTTTATATATAAACAGCGCTATATAGCTGACCTGGCAGTCTCTATAGAGCGCGCA	660
Qy	661	GCCGGCTCTTCTATTAGCTTTGCTCTTATGGCTACATCTGTGTGAGCAGTCGATTGATTC	720
Db	661	GCCGGCTCTTCTATTAGCTTTGCTCTTATGGCTACATCTGTGTGAGCAGTCGATTGATTC	720
Qy	721	AAACAACAATCCGGCGGTTTACGCAAGTCGGAATGAATTTGGCTCATCACTCATTTGTGC	780
Db	721	AAACAACAATCCGGCGGTTTACGCAAGTCGGAATGAATTTGGCTCATCACTCATTTGTGC	780
Qy	781	TGGGCTCTCAGCGCTATTTCCGCTAACCGTGTTCGAATCAGACCCCTCAGAGCCACCGCTC	840
Db	781	TGGGCTCTCAGCGCTATTTCCGCTAACCGTGTTCGAATCAGACCCCTCAGAGCCACCGCTC	840
Qy	841	CAGCAGCCCGTTTACCAACAGCTGAGCTTAAAGAAAGAAAAAACTGTTCAATCACACGCC	900
Db	841	CAGCAGCCCGTTTACCAACAGCTGAGCTTAAAGAAAGAAAAAACTGTTCAATCACACGCC	900
Qy	901	CATCTGAACCGTTTCAACAGCGCCACCGTAAATTTCCGCGCACAGCAAGGCGATATCCGTCA	960
Db	901	CATCTGAACCGTTTCAACAGCGCCACCGTAAATTTCCGCGCACAGCAAGGCGATATCCGTCA	960
Qy	961	TAGCGAGCGATAAATTTCTGATTTCTGCTGCTGCTGCGGCAATTTATCTTTGGGAGGC	1020
Db	961	TAGCGAGCGATAAATTTCTGATTTCTGCTGCTGCTGCGGCAATTTATCTTTGGGAGGC	1020
Qy	1021	GGCGCGGATTGGAGACAGAGCCCAAGGCAACAAGTGGCGTGGAGAAATCAACA	1080
Db	1021	GGCGCGGATTGGAGACAGAGCCCAAGGCAACAAGTGGCGTGGAGAAATCAACA	1080
Qy	1081	AGCGGTGCTTGGCGAGAAGAGAGAGAGAG 1111	
Db	1081	AGCGGTGCTTGGCGAGAAGAGAGAGAGAG 1111	
RESULT 3			
BD263889			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum. 1 (bases 1 to 2687)			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
OS Triticum aestivum (common wheat) PN JP 2002543844-A/1 PD 24-DEC-2002 PF 17-MAY-2000 JP 2000618471 PI MARIE FRANCOISE GAUTIER, TANIA IHOAI, PHILIPPE JOUDRIER PC C12N15/09,A01H5/00,C12N5/10//C12N5/10,C12R1/91,C12N15/00, PC C12N5/00, PC (C12N5/00,C12R1/91) CC Promoter of thioredoxine TaTrxh2 in wheat FH Key Location/Qualifiers FT exon (1112) . (1231) FT intron (1232) . (2203) FT exon (2204) . (2326) FT intron (2327) . (2420) FT exon (2421) . (2558) FT CDS (1112) . (1231) FT CDS (2204) . (2326) FT CDS (2421) . (2558) . 1. .2687			

ORIGIN		Query Match Best Local Similarity 100.0%; Score 1111; DB 6; Length 2687; Matches 1111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1	GAAGTCAGAAAGCCCGTTTCAAGATTGTTGGAGACTCGAAAAAAGAGGGAGCCAGCC	60
Db	1	GAAGTCAGAAAGCCCGTTTCAAGATTGTTGGAGACTCGAAAAAAGAGGGAGCCAGCC	60
Qy	61	AGACGACGGGCGGCATGTGCTGTTCTTGGCGAGGCGTCTAGCTTTGGCAGCCCGCC	120
Db	61	AGACGACGGGCGGCATGTGCTGTTCTTGGCGAGGCGTCTAGCTTTGGCAGCCCGCC	120
Qy	121	CGCTTTTCTTCTTGGTGGGCGGAGCTCCCGAGTTTGGAGCCGCAATTTTTTACAT	180
Db	121	CGCTTTTCTTCTTGGTGGGCGGAGCTCCCGAGTTTGGAGCCGCAATTTTTTACAT	180
Qy	181	TTTATGGCGATGGCGTCAGGCGTTTATCTAGGCGTCTGGAGGGTACATTTGAAGATGTG	240
Db	181	TTTATGGCGATGGCGTCAGGCGTTTATCTAGGCGTCTGGAGGGTACATTTGAAGATGTG	240
Qy	241	CCACCAACTCCAAACCGCAACCCCTGTATCTGAGCATGCTCATGCTCTCTTCAATGCC	300
Db	241	CCACCAACTCCAAACCGCAACCCCTGTATCTGAGCATGCTCATGCTCTCTTCAATGCC	300
Qy	301	TCCCTTTGGTGAGGTCAATGTCCTTGGCGGAGTGGCTTCCGTTTAGAGCAAGTAT	360
Db	301	TCCCTTTGGTGAGGTCAATGTCCTTGGCGGAGTGGCTTCCGTTTAGAGCAAGTAT	360
Qy	361	AATAAGTCTTAGTCAAGTGTATAGATGTTTCCACATCAGCAAAATCTTTAACTGGAGG	420
Db	361	AATAAGTCTTAGTCAAGTGTATAGATGTTTCCACATCAGCAAAATCTTTAACTGGAGG	420
Qy	421	AGAAAGAAAGTAGGAGTGAGAAAGGCGTTCGGCTTCTGTCATTCGTCAGTACGATAGCAAA	480
Db	421	AGAAAGAAAGTAGGAGTGAGAAAGGCGTTCGGCTTCTGTCATTCGTCAGTACGATAGCAAA	480
Qy	481	GCTCCCATGGAATCGAGCCCAACATCAACCCCGCAACATGACTTAAAGGCAAAAGCCAGCCA	540
Db	481	GCTCCCATGGAATCGAGCCCAACATCAACCCCGCAACATGACTTAAAGGCAAAAGCCAGCCA	540
Qy	541	ATCAGTATGCTTTCTCTGATCTTTCTTCATGCAAGCAATTAATACTATAGCTAATCTA	600
Db	541	ATCAGTATGCTTTCTCTGATCTTTCTTCATGCAAGCAATTAATACTATAGCTAATCTA	600
Qy	601	CAGCCAGTTTATATATAAACAGGCTATATAGTCACTGACCTGGCAGTCTTATAGAGCCGCCA	660
Db	601	CAGCCAGTTTATATATAAACAGGCTATATAGTCACTGACCTGGCAGTCTTATAGAGCCGCCA	660
Qy	661	GCGGCTCTCTTATAGCTTTGCTCTTATGGGTACATCTGTGTGAGCAGTCGATTGATTC	720
Db	661	GCGGCTCTCTTATAGCTTTGCTCTTATGGGTACATCTGTGTGAGCAGTCGATTGATTC	720
Qy	721	AAACAACAATCCGGCGGTTTACGCAAGTCGGAATGAATTTCCGCTCATCACTCATTTGTGC	780
Db	721	AAACAACAATCCGGCGGTTTACGCAAGTCGGAATGAATTTCCGCTCATCACTCATTTGTGC	780
Qy	781	TGGGCTCTCAGCGTATTTCCGCTAACCGTGTTCGAATCAGACCCCTCAGAGCCACCGCTC	840
Db	781	TGGGCTCTCAGCGTATTTCCGCTAACCGTGTTCGAATCAGACCCCTCAGAGCCACCGCTC	840
Qy	841	CAGCAGCCCGTTTACCAACAGCTGAGCTTAAAGAAAGAAAAAACTGTTCAATCACACGCC	900
Db	841	CAGCAGCCCGTTTACCAACAGCTGAGCTTAAAGAAAGAAAAAACTGTTCAATCACACGCC	900
Qy	901	CATCTGAACCGTTTCAACAGCGCCACCGTAAATTTCCGCGCACAGCAAGGCGATATCCGTCA	960
Db	901	CATCTGAACCGTTTCAACAGCGCCACCGTAAATTTCCGCGCACAGCAAGGCGATATCCGTCA	960

QY 961 TAGCGAGCGCATAAATTCGTATTCCTGCTCGCTGCCGCAAAATTTATCTTTGGGAGGC 1020  
Db 961 TAGCGAGCGCATAAATTCGTATTCCTGCTCGCTGCCGCAAAATTTATCTTTGGGAGGC 1020  
QY 1021 GGGCCGGGATTTGGAGACAGAGCCCAAGGCAACAAAGTGGCGGTGAGAAATCAACA 1080  
Db 1021 GGGCCGGGATTTGGAGACAGAGCCCAAGGCAACAAAGTGGCGGTGAGAAATCAACA 1080  
QY 1081 AGCGTGCTTCCGAGAGAGAGAGAGAG 1111  
Db 1081 AGCGTGCTTCCGAGAGAGAGAGAGAG 1111

RESULT 4  
AX047681  
LOCUS AX047681 2687 bp DNA linear PAT 15-DEC-2000  
DEFINITION Sequence 1 from Patent WO0070065.  
ACCESSION AX047681  
VERSION AX047681.1 GI:11876716  
KEYWORDS  
SOURCE Triticum aestivum (bread wheat)  
ORGANISM Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Pooideae; Triticeae; Triticum.

REFERENCE  
1 Gautier,M.F., Iborai,T. and Joudrier,P.  
AUTHORS Promoter of thioedoxine tatrzh2 in wheat  
TITLE Patent: WO 0070065-A 1 23-NOV-2000;  
JOURNAL INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (INRA) (FR)  
FEATURES  
source 1. .2687  
/organism="Triticum aestivum"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:4565"  
1112. .1231  
exon 1232. .2203  
intron 2204. .2326  
exon 2327. .2420  
intron 2421. .2558  
exon

ORIGIN  
Query Match 100.0%; Score 1111; DB 6; Length 2687;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GAAGTCAGAGCGCGTTTCAGAAATGTTGGAGGACTCGAAAAAGAGGGGAGCCGAGC 60  
Db 1 GAAGTCAGAGCGCGTTTCAGAAATGTTGGAGGACTCGAAAAAGAGGGGAGCCGAGC 60  
QY 61 AGACGACGGGCGGCATGCTGCTTCCTTGGCAGGCGCTAGCTTTGGCAGCGCGCG 120  
Db 61 AGACGACGGGCGGCATGCTGCTTCCTTGGCAGGCGCTAGCTTTGGCAGCGCGCG 120  
QY 121 CGCTTTTCTCTTGGTGGGCGCGAGCTCCCGAGTTTGAGCGCAATTTTTCAT 180  
Db 121 CGCTTTTCTCTTGGTGGGCGCGAGCTCCCGAGTTTGAGCGCAATTTTTCAT 180  
QY 181 TTTATGGCGATGCGGTTCAGGGGTTTATCTAGCGCTCTGGAGGGTACATTTGAAGATGTG 240  
Db 181 TTTATGGCGATGCGGTTCAGGGGTTTATCTAGCGCTCTGGAGGGTACATTTGAAGATGTG 240  
QY 241 CCACCAATCCAAACCGACACCTGTATCTGAGCATGCCCTCATGCTCTCTTCATGCC 300  
Db 241 CCACCAATCCAAACCGACACCTGTATCTGAGCATGCCCTCATGCTCTCTTCATGCC 300  
QY 301 TCCCTTTGGGTAGTCAATGTCCTTGGCGGCGAGTGGCTTCCCGTTTGAAGCAAGTAT 360  
Db 301 TCCCTTTGGGTAGTCAATGTCCTTGGCGGCGAGTGGCTTCCCGTTTGAAGCAAGTAT 360  
QY 361 AATAAGTCTAGTCAGCTGGCTATAAGATGTTCCACATCAGCAATCTTAAACTGGAGG 420  
Db 361 AATAAGTCTAGTCAGCTGGCTATAAGATGTTCCACATCAGCAATCTTAAACTGGAGG 420

QY 421 AGAAGAAAGTAGGAGTGAGAGGGCGTTCGGCGCTTCGTCAATCGCTAGCGATAGCAAA 480  
Db 421 AGAAGAAAGTAGGAGTGAGAGGGCGTTCGGCGCTTCGTCAATCGCTAGCGATAGCAAA 480  
QY 481 GCTCCCATGGAATCGAGCCCAACATGCAACCCCGCACAAATGACTAAAGGCAACGCCGCA 540  
Db 481 GCTCCCATGGAATCGAGCCCAACATGCAACCCCGCACAAATGACTAAAGGCAACGCCGCA 540  
QY 541 ATCAGTATGCTTTCTCTGCAATCTTTCTTATGCTCAATGCAAGCATTAATACTATAGCTA 600  
Db 541 ATCAGTATGCTTTCTCTGCAATCTTTCTTATGCTCAATGCAAGCATTAATACTATAGCTA 600  
QY 601 CAGCCAGTTTATTAATAACAGGCTATATAGCTGACCTGGCAGTGTATAGAGCCGCA 660  
Db 601 CAGCCAGTTTATTAATAACAGGCTATATAGCTGACCTGGCAGTGTATAGAGCCGCA 660  
QY 661 GCCGGCTCTTTAGCTTTTGTCTTTATGGCTACATCTGTGTGAGCAGTCGATTGATTC 720  
Db 661 GCCGGCTCTTTAGCTTTTGTCTTTATGGCTACATCTGTGTGAGCAGTCGATTGATTC 720  
QY 721 AAACAACAATCCGGGGGTTTCAGCAAGTCGGAATGAATTCGGCTCATCACTCATTTGTG 780  
Db 721 AAACAACAATCCGGGGGTTTCAGCAAGTCGGAATGAATTCGGCTCATCACTCATTTGTG 780  
QY 781 TGGGCTCTCAGCGTATTTTCGGCTAACCGTGTTCGAATCAGACCCCTCAGGAAGCCACGGCTC 840  
Db 781 TGGGCTCTCAGCGTATTTTCGGCTAACCGTGTTCGAATCAGACCCCTCAGGAAGCCACGGCTC 840  
QY 841 CAGCGACCCGTTCCACCGTCAGCCCTAAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 900  
Db 841 CAGCGACCCGTTCCACCGTCAGCCCTAAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 900  
QY 901 CATCTGAAACCGTTCAACAGCGCCACCGTAAATTCGCGCACCCAGCAAGGCGCATATCCGTCA 960  
Db 901 CATCTGAAACCGTTCAACAGCGCCACCGTAAATTCGCGCACCCAGCAAGGCGCATATCCGTCA 960  
QY 961 TAGCGAGCGCATAAATTCGTATTCCTGCTCGCTTCGGGCAATTTATCTTTGGGAGGC 1020  
Db 961 TAGCGAGCGCATAAATTCGTATTCCTGCTCGCTTCGGGCAATTTATCTTTGGGAGGC 1020  
QY 1021 GGGCCGGGATTTGGAGACAGAGCCCAAGGCAACAAAGTGGCGGTGAGAAATCAACA 1080  
Db 1021 GGGCCGGGATTTGGAGACAGAGCCCAAGGCAACAAAGTGGCGGTGAGAAATCAACA 1080  
QY 1081 AGCGTGCTTCCGAGAGAGAGAGAGAG 1111  
Db 1081 AGCGTGCTTCCGAGAGAGAGAGAGAG 1111

RESULT 5  
TAE9762  
LOCUS Triticum aestivum mRNA for thioredoxin H. 596 bp mRNA linear PLN 15-AUG-2001  
DEFINITION Triticum aestivum mRNA for thioredoxin H.  
ACCESSION AJ009762  
VERSION AJ009762.1 GI:4138593  
KEYWORDS thioredoxin H.  
SOURCE Triticum aestivum (bread wheat)  
ORGANISM Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Pooideae; Triticeae; Triticum.

REFERENCE  
1 Serrato,A.J., Crespo,J.L., Florencio,F.J. and Cejudo,F.J.  
AUTHORS Characterization of two thioredoxins h with predominant  
TITLE localization in the nucleus of aleurone and scutellum cells of  
germinating wheat seeds  
JOURNAL Plant Mol. Biol. 46 (3), 361-371 (2001)  
MEDLINE 21380673  
PUBMED 11488482  
REFERENCE 2 (bases 1 to 596)  
AUTHORS Cejudo,F.J.  
TITLE Direct Submission

JOURNAL Submitted (24-JUL-1998) Cejudo F.J., Instituto de Bioquimica Vegetal y Fotosintesis, Universidad de Sevilla y CSIC, Avda Americo Vespucio s/n, 41092-Sevilla, SPAIN

FEATURES source 1. .596 /organism="Triticum aestivum" /mol\_type="mRNA" /cultivar="Chinese Spring" /db\_xref="taxon:4565" 1. .596 /genes="thioredoxin H" 436..596 /gene="thioredoxin H" 462..466 /gene="thioredoxin H"

3'UTR

polyA\_signal

ORIGIN

Query Match 4.9%; Score 54; DB 8; Length 596; Best Local Similarity 100.0%; Pred. No. 28-18; Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1058 AAAGTGGCGGTGAGAAATCAACAGCGTCTTCCGAGAGAGAGAGAGAG 1111

Db 1 AAAGTGGCGGTGAGAAATCAACAGCGTCTTCCGAGAGAGAGAGAGAG 54

RESULT 6 TAE404845 629 bp mRNA linear PLN 15-AUG-2001

LOCUS Triticum aestivum mRNA for thioredoxin h.

DEFINITION AJ404845

ACCESSION AJ404845.1 GI:8980490

VERSION

KEYWORDS thioredoxin H.

SOURCE Triticum aestivum (bread wheat)

ORGANISM Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.

REFERENCE 1 Serrato,A.J., Crespo,J.L., Florencio,F.J. and Cejudo,F.J. Characterization of two thioredoxins h with predominant localization in the nucleus of aleurone and scutellum cells of germinating wheat seeds Plant Mol. Biol. 46 (3), 361-371 (2001)

JOURNAL MEDLINE 21380673

PUBMED 11488482

REFERENCE 2 (bases 1 to 629)

Cejudo,F.J.

Direct Submission

TITLE Submitted (06-JUL-2000) Cejudo F.J., Instituto de Bioquimica Vegetal y Fotosintesis, Universidad de Sevilla, Avda Americo Vespucio s/n, 41092-Sevilla, SPAIN

JOURNAL

FEATURES source 1. .629 /organism="Triticum aestivum" /mol\_type="mRNA" /cultivar="Chinese Spring" /db\_xref="taxon:4565" /tissue\_type="aleurone" /dev\_stage="germinating seed" /country="Spain" 112..489 /note="ORF" /codon\_start=1 /product="thioredoxin h" /protein\_id="CAB96931.1" /db\_xref="GI:8980491" /db\_xref="GO:091DX4" /db\_xref="UniProt/TREMBL:Q9LDX4" /translations="MAASATATAAAGVAGEVISHLSQWTTWQIEEANAAKLVVID FTASWCGPCRIMAFIFADLAKFPAAVFLKVDVDELKSIABQFSVEAMPTFLFMKEGD VKDRVVGAIKEELTNKVGHLAAQ"

3'UTR 490..629 polyA\_signal 516..523

ORIGIN

Query Match 2.9%; Score 32; DB 8; Length 629; Best Local Similarity 100.0%; Pred. No. 6.6e-06; Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1080 AAGCGGTGCTTCCGAGAGAGAGAGAGAGAG 1111

Db 58 AAGCGGTGCTTCCGAGAGAGAGAGAGAGAG 89

RESULT 7 BX901914/c 228676 bp DNA linear HTG 10-OCT-2004

LOCUS BX901914

DEFINITION Danio rerio clone DKEY-27F18, WORKING DRAFT SEQUENCE, 17 unordered pieces.

ACCESSION BX901914

VERSION BX901914.6 GI:54021834

KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_FULLTOP.

SOURCE Danio rerio (zebrafish)

ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.

REFERENCE 1 (bases 1 to 228676) McLaren,S.

AUTHORS Direct Submission

TITLE Submitted (09-OCT-2004) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Oct 10, 2004 this sequence version replaced gi:46194670.

COMMENT ----- Genome Center ----- Center: Wellcome Trust Sanger Institute Center code: SC Web site: http://www.sanger.ac.uk Contact: zfish-help@sanger.ac.uk ----- Project Information ----- Center project name: zK27F18 ----- Summary Statistics ----- Assembly program: XGAP4; version 4.5 Chemistry: Dye-terminator; 100% of reads Consensus quality: 22727 bases at least Q40 Consensus quality: 22391 bases at least Q30 Consensus quality: 224176 bases at least Q20 Insert size: 227076; sum-of-contigs Insert size: 222615; 5.7% error; agarose-fp Quality coverage: 6.94x in Q20 bases; sum-of-contigs Quality coverage: 7.21x in Q20 bases; agarose-fp -----

\* NOTE: This is a 'working draft' sequence. It currently consists of 17 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 10555: contig of 10555 bp in length

10556: gap of 100 bp

10656: contig of 10079 bp in length

20735: gap of 100 bp

20835: contig of 8597 bp in length

29432: gap of 100 bp

29532: contig of 17847 bp in length

47379: gap of 100 bp

47479: contig of 2105 bp in length

49584: gap of 100 bp

49684: contig of 25600 bp in length

75284: gap of 100 bp

75384: contig of 18206 bp in length

93590: gap of 100 bp

```

* 93690 96129: contig of 2440 bp in length
* 96130 96229: gap of 100 bp
* 96230 109725: contig of 13496 bp in length
* 109726 109825: gap of 100 bp
* 109826 124219: contig of 14394 bp in length
* 124220 124319: gap of 100 bp
* 124320 129512: contig of 5193 bp in length
* 129513 129612: gap of 100 bp
* 129613 134036: contig of 4424 bp in length
* 134037 134136: gap of 100 bp
* 134137 151328: contig of 17192 bp in length
* 151329 151428: gap of 100 bp
* 151429 167242: contig of 15814 bp in length
* 167243 167342: gap of 100 bp
* 167343 179999: contig of 12657 bp in length
* 180000 180099: gap of 100 bp
* 180100 190992: contig of 10893 bp in length
* 190993 191092: gap of 100 bp
* 191093 228676: contig of 37584 bp in length.
FEATURES
    source
        1..228676
            /organism="Danio rerio"
            /mol_type="genomic DNA"
            /db_xref="taxon:7955"
            /clone="DKEY-27F18"
            /clone_lib="DanioKey"
            1..10555
                /note="assembly fragment:00271"
                fragment_chain:1
                10656..20734
                    /note="assembly fragment:00142"
                    fragment_chain:1
                    20835..29431
                        /note="assembly fragment:00056"
                        fragment_chain:1
                        29532..47378
                            /note="assembly fragment:01109"
                            fragment_chain:1
                            47479..49583
                                /note="assembly fragment:02902"
                                fragment_chain:1
                                49684..75283
                                    /note="assembly fragment:02010"
                                    fragment_chain:2
                                    75384..93589
                                        /note="assembly fragment:01546"
                                        fragment_chain:2
                                        93690..96129
                                            /note="assembly fragment:02882"
                                            fragment_chain:2
                                            96230..109725
                                                /note="assembly fragment:00733"
                                                fragment_chain:2
                                                109826..124219
                                                    /note="assembly fragment:00914"
                                                    fragment_chain:2
                                                    124320..129512
                                                        /note="assembly fragment:02928"
                                                        fragment_chain:3
                                                        129613..134036
                                                            /note="assembly fragment:00001"
                                                            fragment_chain:3
                                                            134137..151328
                                                                /note="assembly fragment:01324"
                                                                fragment_chain:3
                                                                151429..167242
                                                                    /note="assembly fragment:01775"
                                                                    fragment_chain:3
                                                                    167343..179999
                                                                        /note="assembly fragment:00411"
                                                                        fragment_chain:3
                                                                        180100..190992
                                                                            /note="assembly fragment:00565.0"

```

```

misc_feature 191093..228676
              /note="assembly_fragment:02360"
ORIGIN
Query Match      2.3%   Score 26;   DB 2;   Length 228676;
Best Local Similarity 100.0%;   Pred. No. 0.018;   0;   Gaps 0;
Matches 26;   Conservative 0;   Mismatches 0;   Indels 0;
QY      867  AAAAAAGAAAAAACTGTTCAAT 892
          |||||
Db      122236  AAAAAAGAAAAAACTGTTCAAT 122211

RESULT 8
AC021823/c
LOCUS      146015 bp   DNA   linear   HTG 04-APR-2000
DEFINITION Homo sapiens chromosome 2 clone RP11-169P13 map 2, WORKING DRAFT
SEQUENCE, 22 unordered pieces.
AC021823
VERSION    AC021823.3   GI:7408019
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1  (bases 1 to 146015)
            Birren,B., Linton,L., Nusbaum,C. and Lander,E.
            Homo sapiens chromosome 2, clone RP11-169P13
            Unpublished
REFERENCE  2  (bases 1 to 146015)
            Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavskiy,L., Bouckgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Collangelo,M., Collins,S., Collimore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenesfor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierfe,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
            Direct Submission
            Submitted (20-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            On Apr 4, 2000 this sequence version replaced GI:6937782.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L5939
            Center clone name: 169_P_13
            ----- Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 135544 bases at least Q40
            Consensus quality: 140389 bases at least Q30
            Consensus quality: 142539 bases at least Q20
            Insert size: 153000; agarose-fp
            Insert size: 143915; sum-of-contigs
            Quality coverage: 4.1 in Q20 bases; agarose-fp
            Quality coverage: 4.4 in Q20 bases; sum-of-contigs

```

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 22 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 40: contig of 40 bp in length  
\* 41 140: gap of 100 bp  
\* 141 1665: contig of 1525 bp in length  
\* 1666 1765: gap of 100 bp  
\* 1766 3194: contig of 1429 bp in length  
\* 3195 3294: gap of 100 bp  
\* 3295 5598: contig of 2304 bp in length  
\* 5599 5698: gap of 100 bp  
\* 5699 10030: contig of 4332 bp in length  
\* 10031 10130: gap of 100 bp  
\* 10131 13650: contig of 3520 bp in length  
\* 13651 13750: gap of 100 bp  
\* 13751 18728: contig of 4978 bp in length  
\* 18729 18828: gap of 100 bp  
\* 18829 21761: contig of 2932 bp in length  
\* 21761 21860: gap of 100 bp  
\* 21861 25279: contig of 3419 bp in length  
\* 25280 25379: gap of 100 bp  
\* 25380 29185: contig of 3806 bp in length  
\* 29186 29285: gap of 100 bp  
\* 29286 33428: contig of 4143 bp in length  
\* 33429 33528: gap of 100 bp  
\* 33529 39961: contig of 6433 bp in length  
\* 39962 40061: gap of 100 bp  
\* 40062 43385: contig of 3304 bp in length  
\* 43386 43485: gap of 100 bp  
\* 43486 48258: contig of 4793 bp in length  
\* 48259 48358: gap of 100 bp  
\* 48359 54546: contig of 6188 bp in length  
\* 54547 54646: gap of 100 bp  
\* 54647 65171: contig of 10525 bp in length  
\* 65172 65271: gap of 100 bp  
\* 65272 78708: contig of 13437 bp in length  
\* 78709 78808: gap of 100 bp  
\* 78809 90249: contig of 11441 bp in length  
\* 90250 90349: gap of 100 bp  
\* 90350 102852: contig of 12503 bp in length  
\* 102853 102952: gap of 100 bp  
\* 102953 114466: contig of 11514 bp in length  
\* 114467 114566: gap of 100 bp  
\* 114567 127728: contig of 13162 bp in length  
\* 127729 127829: gap of 100 bp  
\* 127829 146015: contig of 18187 bp in length.

FEATURES

source  
1. .146015  
/organism="Homo sapiens"  
/mol type="genomic DNA"  
/db xref="taxon:9606"  
/chromosome="2"  
/map="2"  
/clone="RP11-169P13"  
/clone lib="RP11-11 Human Male BAC"  
misc\_feature 1. .40  
/note="assembly\_fragment  
clone\_end:17  
vector\_side:right"  
misc\_feature 141. .1665  
/note="assembly\_fragment"  
misc\_feature 1766. .3194  
/note="assembly\_fragment"  
misc\_feature 3295. .5598  
/note="assembly\_fragment"  
misc\_feature 5699. .10030  
/note="assembly\_fragment"

misc\_feature 10131. .13650  
/note="assembly\_fragment"  
misc\_feature 13751. .18728  
/note="assembly\_fragment"  
misc\_feature 18829. .21760  
/note="assembly\_fragment"  
misc\_feature 21861. .25279  
/note="assembly\_fragment"  
misc\_feature 25380. .29185  
/note="assembly\_fragment"  
misc\_feature 29286. .33428  
/note="assembly\_fragment"  
misc\_feature 33529. .39961  
/note="assembly\_fragment"  
misc\_feature 40062. .43365  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:right"  
misc\_feature 43466. .48258  
/note="assembly\_fragment"  
misc\_feature 48359. .54546  
/note="assembly\_fragment"  
misc\_feature 54647. .65171  
/note="assembly\_fragment"  
misc\_feature 65272. .78708  
/note="assembly\_fragment"  
misc\_feature 78809. .90249  
/note="assembly\_fragment"  
misc\_feature 90350. .102852  
/note="assembly\_fragment"  
misc\_feature 102953. .114466  
/note="assembly\_fragment"  
misc\_feature 114567. .127728  
/note="assembly\_fragment"  
misc\_feature 127829. .146015  
/note="assembly\_fragment"

ORIGIN

Query Match 2.2%; Score 24; DB 2; Length 146015;  
Best Local Similarity 100.0%; Pred. No. 0.24; Mismatches 0; Indels 0; Gaps 0;  
Matches 24; Conservative 0;

QY 865 CTAAGAAAAAGAAAAAACTGTT 888  
|||||  
Db 112269 CTAAGAAAAAGAAAAAACTGTT 112246

RESULT 9

AC092642  
LOCUS AC092642 152251 bp DNA linear PRI 01-MAR-2002  
DEFINITION Homo sapiens BAC clone RP11-391P1 from 2, complete sequence.  
ACCESSION AC092642 AC023761  
VERSION AC092642.2 GI:15638834  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 2 (bases 1 to 152251)  
AUTHORS Mulvaney,E., Abbott,A., Dixon,R., Dignan,G. and Phillips,A.  
TITLE The sequence of Homo sapiens BAC clone RP11-391P1  
JOURNAL Unpublished (2001)  
MEDLINE 99063792  
PUBMED 9847074  
REFERENCE 2 (bases 1 to 152251)  
AUTHORS Mulvaney,E., Abbott,A., Dixon,R., Dignan,G. and Phillips,A.  
TITLE The sequence of Homo sapiens BAC clone RP11-391P1  
JOURNAL Unpublished (2001)  
REFERENCE 3 (bases 1 to 152251)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (19-JUL-2001) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA  
 4 (bases 1 to 152251)  
 Waterston, R.H.  
 Direct Submission  
 Submitted (18-SEP-2001) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 5 (bases 1 to 152251)  
 Waterston, R.  
 Direct Submission  
 Submitted (01-MAR-2002) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On Sep 18, 2001 this sequence version replaced gi:14916226.  
 ----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc>  
 Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
 ----- Summary Statistics  
 Center project name: H\_NH0391P01  
 Drafting Center: WtBR  
 -----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Oseogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>  
 VECTOR: pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-544E11, 2000 bp overlap; the clone sequenced to the right is RP11-710J17, 2000 bp overlap. Actual end of this clone is at base position 19440 of RP11-710J17.

Data from AC062033, AC061960, and AC021823 was used to finish this clone, AC023761. Polymorphisms have been identified between AC062033, AC061960, and AC023761.

The sequence of AC023761 has been incorporated into AC092642.

#### FEATURES source

1..152251  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 /map="2"  
 /clone="RP11-391P1"  
 /clone.lib="RPCI-11"  
 4..169  
 repeat\_region  
 /rpt\_family="MaLR"

repeat\_region  
 206..495  
 /rpt\_family="MaLR"  
 repeat\_region  
 563..718  
 /rpt\_family="L2"  
 repeat\_region  
 1151..1293  
 /rpt\_family="L2"  
 repeat\_region  
 1524..1624  
 /rpt\_family="L1"  
 repeat\_region  
 1622..1643  
 /rpt\_family="AT\_rich"  
 repeat\_region  
 1628..1694  
 /rpt\_family="Alu"  
 repeat\_region  
 1695..2139  
 /rpt\_family="L1"  
 repeat\_region  
 1794..1837  
 /rpt\_family="L1"  
 repeat\_region  
 1794..1837  
 /rpt\_family="L1"  
 repeat\_region  
 2149..2433  
 /rpt\_family="MaLR"  
 repeat\_region  
 2449..3161  
 /rpt\_family="L1"  
 repeat\_region  
 3162..3456  
 /rpt\_family="Alu"  
 repeat\_region  
 3325..3365  
 /rpt\_family="L1"  
 repeat\_region  
 3457..3758  
 /rpt\_family="L1"  
 repeat\_region  
 3759..4066  
 /rpt\_family="Alu"  
 repeat\_region  
 4067..4492  
 /rpt\_family="L1"  
 repeat\_region  
 4493..4779  
 /rpt\_family="Alu"  
 repeat\_region  
 4780..6269  
 /rpt\_family="L1"  
 repeat\_region  
 6300..6363  
 /rpt\_family="L1"  
 repeat\_region  
 6367..6448  
 /rpt\_family="ERVL"  
 repeat\_region  
 6470..6731  
 /rpt\_family="ERVL"  
 repeat\_region  
 6737..6876  
 /rpt\_family="L1"  
 repeat\_region  
 7608..7643  
 /rpt\_family="CA)n"  
 repeat\_region  
 7956..8394  
 /rpt\_family="MaLR"  
 repeat\_region  
 9284..9382  
 /rpt\_family="L1"  
 repeat\_region  
 9480..9537  
 /rpt\_family="AT\_rich"  
 repeat\_region  
 9534..9825  
 /rpt\_family="Alu"  
 9934  
 misc\_feature  
 /note="match to EST AI038041 (NID:g3277235) ox29h03.x1"  
 misc\_feature  
 10044..10443  
 /note="match to EST AI038041 (NID:g3277235) ox29h03.x1"  
 repeat\_region  
 10054..10075  
 /rpt\_family="AT\_rich"  
 misc\_feature  
 10780..11226  
 /note="match to EST T91378 (NID:g723291) yd53a07.s1"  
 misc\_feature  
 10799..10943  
 /note="match to EST H81620 (NID:g1059709) ys66d09.s1"  
 repeat\_region  
 10853..10901  
 /rpt\_family="L1"  
 misc\_feature  
 11080..11325  
 /note="match to EST H81620 (NID:g1059709) ys66d09.s1"  
 misc\_feature  
 11246..11379  
 /note="match to EST T84526 (NID:g712878) yd53a07.r1"  
 repeat\_region  
 11828..11945  
 /rpt\_family="L2"  
 misc\_feature  
 12091..12249  
 /note="match to EST H81619 (NID:g1059708) ys66d09.r1"  
 misc\_feature  
 12098..12249



```
/note="match to EST T84526 (NID:g712878) yd53a07.r1"
repeat_region 12315..12336
/rpt_family="AT_rich"
repeat_region 15388..15522
/rpt_family="L2"
repeat_region 15624..15788
/rpt_family="(TC)n"
repeat_region 15831..15898
/rpt_family="(TC)n"
repeat_region 15934..16232
/rpt_family="Alu"
repeat_region 18502..18571
/rpt_family="L2"
repeat_region 19347..19607
/rpt_family="Alu"
repeat_region 20299..20328
/rpt_family="AT_rich"
repeat_region 20300..20413
/rpt_family="L1"
repeat_region 20821..21353
/rpt_family="L2"
repeat_region 21356..21468
/rpt_family="L1"
repeat_region 21449..21505
/rpt_family="AT_rich"
repeat_region 21504..21678
/rpt_family="MER1_type"
repeat_region 22507..22594
/rpt_family="Achobo"
repeat_region 22644..23228

Query Match 2.2%; Score 24; DB 9; Length 152251;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 865 CTAAGAAAAAGAAAAAACTGTT 888
|||||
Db 26051 CTAAGAAAAAGAAAAAACTGTT 26074

RESULT 10
BX927299
LOCUS BX927299 155054 bp DNA linear HTG 08-OCT-2004
DEFINITION Danio rerio clone DKEY-146H10, WORKING DRAFT SEQUENCE, 7 unordered
pieces.
ACCESSION BX927299
VERSION BX927299.5 GI:54019870
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.

REFERENCE 1 (bases 1 to 155054)
AUTHORS Heath, P.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zf1sh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Oct 9, 2004 this sequence version replaced gi:52313301.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zf1sh-help@sanger.ac.uk
----- Project Information
Center project name: zK146H10
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 153838 bases at least Q40
Consensus quality: 153985 bases at least Q30
Consensus quality: 154096 bases at least Q20
```

```
Insert size: 154454; sum-of-contigs
Insert size: 158561; 1.9% error; agarose-fp
Quality coverage: 11.18x in Q20 bases; sum-of-contigs Quality
coverage: 10.90x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 33656: contig of 33656 bp in length
* 33657 33756: gap of 100 bp
* 33757 51099: contig of 17343 bp in length
* 51100 51199: gap of 100 bp
* 51200 60096: contig of 8897 bp in length
* 60097 60196: gap of 100 bp
* 60197 85868: contig of 25672 bp in length
* 85869 85968: gap of 100 bp
* 85969 116591: contig of 30623 bp in length
* 116592 116691: gap of 100 bp
* 116692 147668: contig of 30977 bp in length
* 147669 147769: gap of 100 bp
* 147769 155054: contig of 7286 bp in length.
FEATURES
Location/Qualifiers
source 1..155054
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-146H10"
/clone_lib="DanioKey"
1..33656
/note="assembly_fragment:00001
fragment_chain:1"
misc_feature 33757..51099
/note="assembly_fragment:02384
fragment_chain:1"
misc_feature 51200..60096
/note="assembly_fragment:02382
fragment_chain:2"
misc_feature 60197..85868
/note="assembly_fragment:01606
fragment_chain:2"
misc_feature 85969..116591
/note="assembly_fragment:00526.0"
misc_feature 116692..147668
/note="assembly_fragment:01071"
misc_feature 147769..155054
/note="assembly_fragment:02238"

ORIGIN
Query Match 2.2%; Score 24; DB 2; Length 155054;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 868 AAAAAAGAAAAAACTGTTCAA 891
|||||
Db 92426 AAAAAAGAAAAAACTGTTCAA 92449

RESULT 11
BX001022/c
LOCUS BX001022 160804 bp DNA linear VRT 11-OCT-2003
DEFINITION Zebrafish DNA sequence from clone CH211-243019, complete sequence.
ACCESSION BX001022
VERSION BX001022.9 GI:37652295
KEYWORDS HTG.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
```

Cypriniformes; Cyprinidae; Danio.  
1. (bases 1 to 160804)  
Giselle.H.  
Direct Submission  
Submitted (11-OCT-2003) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Oct 13, 2003 this sequence version replaced gi:35209032.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: zfish-help@sanger.ac.uk  
-----  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest, except on the rare  
occasion of the clone being a YAC.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information  
on the WORMPEP database can be found at  
<http://www.sanger.ac.uk/projects/C.elegans/wormpep> Clone-derived  
Zebrafish pUC subclones occasionally display inconsistency over the  
length of mononucleotide A/T runs and conserved TA repeats. Where  
this is found the longest good quality representation will be  
submitted.  
Repeat names beginning 'Dr' were identified by the Recon repeat  
discovery system (Zhiron Bao and Sean Eddy, submitted), and those  
beginning 'dr' were identified by Rick Waterman (Stephen Johnson  
lab, WashU). For further information see  
<http://www.sanger.ac.uk/Projects/D.rerio/fishmask.shtml>  
CH211-243019 is from a CHORI-211 BAC library  
VECTOR: pTARBAC2.1.  
-----  
Location/Qualifiers  
1..160804  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clones="CH211-243019"  
/clone\_lib="CHORI-211"  
-----  
ORIGIN  
Query Match 2.2%; Score 24; DB 5; Length 160804;  
Best Local Similarity 100.0%; Pred. No. 0.24;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
-----  
Qy 868 AAAAAAGAAAAAACTGTTCAA 891  
|||||  
Db 44746 AAAAAAGAAAAAACTGTTCAA 44723  
-----  
FEATURES  
source  
1..160804  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clones="CH211-243019"  
/clone\_lib="CHORI-211"  
-----  
ORIGIN  
Query Match 2.2%; Score 24; DB 5; Length 160804;  
Best Local Similarity 100.0%; Pred. No. 0.24;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
-----  
Qy 868 AAAAAAGAAAAAACTGTTCAA 891  
|||||  
Db 44746 AAAAAAGAAAAAACTGTTCAA 44723  
-----  
RESULT 12  
BX927400/c  
LOCUS BX927400 185217 bp DNA linear HTG 10-OCT-2004  
DEFINITION Danio rerio clone CH211-195K18, WORKING DRAFT SEQUENCE.  
ACCESSION BX927400  
VERSION BX927400.6 GI:51965253  
KEYWORDS HTG; HTGS\_PHASE2; HTGS\_ACTIVEPIN; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE Danio rerio (zebrafish)  
ORGANISM Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Danio.  
1. (bases 1 to 185217)  
Ellwood.M.  
Direct Submission  
Submitted (08-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Sep 9, 2004 this sequence version replaced gi:51592026.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: zfish-help@sanger.ac.uk  
----- Project Information  
Center project name: zC195K18  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Chemistry: Dye-terminator; 100% of reads  
Consensus quality: 184890 bases at least Q40  
Consensus quality: 185061 bases at least Q30  
Consensus quality: 185152 bases at least Q20  
Insert size: 185217; sum-of-contigs  
Insert size: 210440; 9.8% error; agarose-fp  
Quality coverage: 9.78x in Q20 bases; sum-of-contigs Quality  
coverage: 8.61x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 185217: contig of 185217 bp in length.  
-----  
Location/Qualifiers  
1..185217  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clone="CH211-195K18"  
/clone\_lib="CHORI-211"  
-----  
misc\_feature  
1..185217  
/note="assembly\_fragment:02923  
clone\_end:SP6  
vector\_side:right  
clone\_end:T7  
vector\_side:left"  
-----  
ORIGIN  
Query Match 2.2%; Score 24; DB 2; Length 185217;  
Best Local Similarity 100.0%; Pred. No. 0.24;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
-----  
Qy 868 AAAAAAGAAAAAACTGTTCAA 891  
|||||  
Db 8803 AAAAAAGAAAAAACTGTTCAA 8780  
-----  
RESULT 13  
BX649502/c  
LOCUS BX649502 247387 bp DNA linear VRT 06-DEC-2003  
DEFINITION Zebrafish DNA sequence from clone DKEY-204F11 in linkage group 3,  
complete sequence.  
ACCESSION BX649502  
VERSION BX649502.4 GI:39540484  
KEYWORDS HTG.  
SOURCE Danio rerio (zebrafish)  
ORGANISM Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.

REFERENCE 1 (bases 1 to 247387)  
 AUTHORS Whitehead,S.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-DEC-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Dec 6, 2003 this sequence version replaced gi:38201304.  
 COMMENT ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: zfish-help@sanger.ac.uk  
 -----  
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.  
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C\_elegans/wormpep  
 Clone-derived Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.  
 Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhiron Bao and Sean Eddy, submitted), and those beginning 'drr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.sanger.ac.uk/projects/D\_rerio/fishmask.shtml DKEY-204F11 is from a Zebrafish BAC library VECTOR: pIndigoBAC-5.  
 FEATURES  
 source  
 1..247387  
 /organism="Danio rerio"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7955"  
 /clone="DKEY-204F11"  
 /clone\_lib="DanioKey"  
 ORIGIN  
 Query Match 2.2%; Score 24; DB 5; Length 247387;  
 Best Local Similarity 100.0%; Pred.No. 0.24;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 867 AAAAAAGAAAAAACTCTTCA 890  
 |||||  
 Db 64795 AAAAAAGAAAAAACTCTTCA 64772  
 RESULT 14  
 BX927111/c  
 LOCUS BX927111 263169 bp DNA linear HTG 05-APR-2004  
 DEFINITION Danio rerio clone DKEY-54K13, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 12 unordered pieces.  
 ACCESSION BX927111  
 VERSION BX927111.5 GI:46237716  
 KEYWORDS HTG; HTGS PHASE1.  
 SOURCE Danio rerio (zebrafish)  
 ORGANISM Danio rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Danio.  
 1 (bases 1 to 263169)  
 Sims,S.  
 Direct Submission  
 Submitted (04-APR-2004) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Apr 6, 2004 this sequence version replaced gi:46200415.  
 COMMENT ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: zfish-help@sanger.ac.uk  
 ----- Project Information  
 Center project name: ZK54K13  
 ----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Chemistry: Dye-terminator; 100% of reads  
 Consensus quality: 260222 bases at least Q40  
 Consensus quality: 260911 bases at least Q30  
 Consensus quality: 261370 bases at least Q20  
 Insert size: 262069; sum-of-contigs  
 Insert size: 249466; 4.6% error; agarose-fp  
 Quality coverage: 8.8ix in Q20 bases; sum-of-contigs Quality coverage: 9.36x in Q20 bases; agarose-fp  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently consists of 12 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  
 \* 1 8720: contig of 8720 bp in length  
 \* 8721 8820: gap of 100 bp  
 \* 8821 34199: contig of 25379 bp in length  
 \* 34299: gap of 100 bp  
 \* 34300 61537: contig of 27238 bp in length  
 \* 61538 61637: gap of 100 bp  
 \* 61638 82341: contig of 20704 bp in length  
 \* 82342 82441: gap of 100 bp  
 \* 82442 113587: contig of 31146 bp in length  
 \* 113588 113687: gap of 100 bp  
 \* 113688 117389: contig of 3702 bp in length  
 \* 117390 117489: gap of 100 bp  
 \* 117490 169953: contig of 52464 bp in length  
 \* 169954 170053: gap of 100 bp  
 \* 170054 191893: contig of 21840 bp in length  
 \* 191894 191993: gap of 100 bp  
 \* 191994 232102: contig of 40109 bp in length  
 \* 232103 232202: gap of 100 bp  
 \* 232203 235712: contig of 3510 bp in length  
 \* 235713 235812: gap of 100 bp  
 \* 235813 256314: contig of 20502 bp in length  
 \* 256315 256414: gap of 100 bp  
 \* 256415 263169: contig of 6755 bp in length.  
 FEATURES  
 Location/Qualifiers  
 1..263169  
 /organism="Danio rerio"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7955"  
 /clone="DKEY-54K13"  
 /clone\_lib="DanioKey"  
 misc\_feature  
 1..8720  
 /note="assembly\_fragment:00210  
 fragment\_chain:1"  
 8821..34199  
 /note="assembly\_fragment:01606  
 fragment\_chain:1"  
 misc\_feature  
 34300..61537  
 /note="assembly\_fragment:01245  
 fragment\_chain:1"

```

misc_feature 61638. .82341
/note="assembly fragment:00629
fragment_chain:1"
misc_feature 82442. .113587
/note="assembly fragment:01994
fragment_chain:1"
misc_feature 113688. .117389
/note="assembly fragment:00023
fragment_chain:1"
misc_feature 117490. .169953
/note="assembly fragment:03001
fragment_chain:1"
misc_feature 170054. .191893
/note="assembly fragment:00915
fragment_chain:1"
misc_feature 191994. .232102
/note="assembly fragment:02450
fragment_chain:1"
misc_feature 232203. .235712
/note="assembly fragment:00058
fragment_chain:1"
misc_feature 235813. .256314
/note="assembly fragment:00359
fragment_chain:1"
misc_feature 256415. .263169
/note="assembly fragment:00104
fragment_chain:1"

ORIGIN
Query Match 2.2%; Score 24; DB 2; Length 263169;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 867 AAAAAAAAAAAAAAAAAACTGTTC A 890
Db 251704 AAAAAAAAAAAAAAAAAACTGTTC A 251681

RESULT 15
AC128496/c
LOCUS AC128496 291544 bp DNA linear HTG 19-NOV-2002
DEFINITION Rattus norvegicus clone CH230-444D9, WORKING DRAFT SEQUENCE, 3
unordered pieces.
ACCESSION AC128496
VERSION AC128496.3 GI:25085139
KEYWORDS HTG: HTGS PHASE1: HTGS DRAFT: HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 291544)
Muzny,D.Mario., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alibrooke,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biewald,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Cesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,X., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,F., Garza,M.,
Gebrgeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,K., Jolivet,A.,

```

```

Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemeleh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Prannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,K., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 291544)
Worley,K.C.
Direct Submission
Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 291544)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 19, 2002 this sequence version replaced gi:23907748.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GYSW
Center clone name: CH230-444D9
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 187201 bases at least Q40
Consensus quality: 198849 bases at least Q30
Consensus quality: 191486 bases at least Q20
Estimated insert size: 193264; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

```

\* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 288506: contig of 288506 bp in length  
\* 288507 288606: gap of unknown length  
\* 288607 289949: contig of 1343 bp in length  
\* 289950 290049: gap of unknown length  
\* 290050 291544: contig of 1495 bp in length.

FEATURES

source  
1. .291544  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-444D9"  
misc\_feature  
1. .1841  
/note="wgs end\_extension  
clone\_end:T7"  
misc\_feature  
6777..7677  
/note="clone\_boundary  
clone\_end:T7  
site:  
end sequence:B2198832"  
complement(187560..188645)  
/note="clone\_boundary  
clone\_end:Sp6  
site:  
end sequence:B2198833"  
189481..192835  
/note="wgs end\_extension  
clone\_end:Sp6"  
195009..196648  
/note="wgs end\_extension  
clone\_end:Sp6"

ORIGIN

Query Match 2.2%; Score 24; DB 2; Length 291544;  
Best Local Similarity 100.0%; Pred. No. 0.24;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 867 AAAAAAAAAAAAAAAAAAACTCTTCA 890  
|||||  
Db 130151 AAAAAAAAAAAAAAAAAAACTCTTCA 130128

Search completed: August 30, 2005, 18:23:47  
Job time : 5165 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 30, 2005, 13:50:45 ; Search time 701 Seconds  
(without alignments)  
9382.072 Million cell updates/sec

Title: US-09-979-549-2  
Perfect score: 1111  
Sequence: 1 gaagtcagaagccgttcag.....ccgagaagagagagagag 1111

Scoring table: OLIGO NUC  
Gapop\_60.0 , Gapext 60.0

Searched: 4390206 seqs, 2959870667 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002as:\*  
7: Geneseqn2002bs:\*  
8: Geneseqn2003as:\*  
9: Geneseqn2003bs:\*  
10: Geneseqn2003cs:\*  
11: Geneseqn2003ds:\*  
12: Geneseqn2004as:\*  
13: Geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1111	100.0	2687	5 AAC841132	Aac841132 Wheat TaT
2	22	2.0	141	6 ABL86452	Ab186452 Human ova
3	22	2.0	144	6 ABL86187	Ab186187 Human ova
4	22	2.0	154	6 ABL86321	Ab186321 Human ova
5	22	2.0	213	6 ABL86948	Ab186948 Human ova
6	22	2.0	1074	5 ADL62118	Ad162118 Human ova
7	22	2.0	24935	4 ABL10390	Ab110390 Drosophila
8	21	1.9	377	2 AAV86132	Aav86132 EST clone
9	21	1.9	403	3 AAC24900	Aac24900 Human sec
10	21	1.9	621	4 AAH71551	Aah71551 Human cer
11	21	1.9	1097	3 AAF16195	Aaf16195 Human pro
12	21	1.9	2000	8 ADA72818	Ada72818 Rice gene
13	21	1.9	2170	4 AAH72880	Aah72880 Human cer
14	21	1.9	2170	5 ABV22589	Abv22589 Human pro
15	21	1.9	2170	5 ABV25695	Abv25695 Human pro
16	21	1.9	2170	5 ABV28409	Abv28409 Human pro
17	21	1.9	2170	5 ABV24749	Abv24749 Human pro
18	21	1.9	3191	3 AAA16647	Aaa16647 Human sec
19	21	1.9	3618	6 ABL56484	Ab156484 Nucleotid
20	21	1.9	4522	3 AAC76440	Aac76440 Human ORF

C 21	21	1.9	4740	5 ABV21724	Abv21724 Human pro
C 22	21	1.9	4740	5 ABV27546	Abv27546 Human pro
C 23	21	1.9	10351	6 ABK88932	Abk88932 Human Cbl
C 24	21	1.9	80815	13 ABD33381	Abd33381 Human can
25	21	1.9	254396	12 ADQ97557	Adq97557 Human can
26	20	1.8	162	6 ABZ08670	Abz08670 Human leu
C 27	20	1.8	371	6 ABK64451	Abk64451 Human ben
28	20	1.8	375	5 ABV48688	Abv48688 Human pro
29	20	1.8	419	4 AA185788	Aa185788 Human pol
30	20	1.8	422	6 ABQ58941	Abq58941 Human col
C 31	20	1.8	461	2 AAV89657	Aav89657 EST clone
32	20	1.8	462	9 ACH22705	Ach22705 Human adu
33	20	1.8	555	4 AAS36746	Aas36746 Human car
34	20	1.8	555	10 ADE47440	Ade47440 Human car
35	20	1.8	555	13 ADJ08858	Adj08858 Human car
C 36	20	1.8	603	4 AAK52328	Aak52328 Human pol
C 37	20	1.8	722	6 ADB87555	Adb87555 Plaemid p
C 38	20	1.8	860	4 AAH03455	Aah03455 Human cDN
C 39	20	1.8	870	3 AAA08590	Aaa08590 Human cyt
C 40	20	1.8	1952	12 ADI42697	Adi42697 Plant tra
C 41	20	1.8	1952	12 ADO02953	Ado02953 Soybean o
42	20	1.8	2143	11 ACN92902	Acn92902 Breast ca
43	20	1.8	2559	13 ACN38938	Acn38938 Tumour-as
44	20	1.8	2577	2 AAX87624	Aax87624 Set* spli
45	20	1.8	2577	6 ABS73234	Abs73234 DNA encod

ALIGNMENTS

```
RESULT 1
AAC84132
ID AAC841132 standard; DNA; 2687 BP.
XX
XX AAC841132;
DT 09-APR-2001 (first entry)
XX
DE Wheat TaTrxh2 gene.
KW Wheat; TaTrxh2; thioredoxin; promoter; gene expression; transgenic plant;
KW monocotyledon; plant cell; seed; amylaceous albumen; da.
XX
OS Triticum aestivum.
XX
FH Key Location/Qualifiers
FT promoter 1..1111
FT /tag= a
FT /note= "promoter region is specifically claimed"
FT misc_signal 550..558
FT /tag= b
FT /label= Gibberellic acid response element
FT /function= "regulates gene expression in response to
FT gibberellic acid"
FT misc_signal 561..569
FT /tag= c
FT /label= Gibberellic acid response element
FT /function= "regulates gene expression in response to
FT gibberellic acid"
FT protein_bind 699..708
FT /tag= d
FT /bound_moiety= "GCNA-like protein"
FT /label= GCNA-like_box
FT protein_bind 860
FT /tag= e
FT /bound_moiety= "leucine zipper proteins"
FT /label= bzlp_motif
FT misc_signal 867..883
FT /tag= f
FT /function= "possible regulator of gene expression in
FT response to abscisic acid"
FT /note= "present in thioredoxin h gene promoter sequences
FT from tobacco and rice"
```





QY 961 TAGCGAGCGCATAAATTCGATTCCTGCTGCTGCGGACAAATTTATCTTTGGGAGGC 1020  
DB 961 TAGCGAGCGCATAAATTCGATTCCTGCTGCTGCGGACAAATTTATCTTTGGGAGGC 1020  
QY 1021 GGCGCGGATTTGAGACAGAGCCACAGGCAACAAGTCGCGTGAGAAATCAACA.1080  
DB 1021 GGCGCGGATTTGAGACAGAGCCACAGGCAACAAGTCGCGTGAGAAATCAACA.1080  
QY 1081 AGCGGTGCTTGGCCGAGAGAGAGAGAGAG 1111  
DB 1081 AGCGGTGCTTGGCCGAGAGAGAGAGAGAG 1111

RESULT 2  
ABL86452  
ID ABL86452 standard; cDNA; 141 BP.  
XX  
AC ABL86452;  
XX  
17-MAY-2002 (first entry)  
DE Human ovarian cancer related cDNA clone SEQ ID NO:9430.  
XX  
XX Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.  
XX  
XX Homo sapiens.  
XX  
PN W0200192581-A2.  
XX  
PD 06-DEC-2001.  
XX  
XX 29-MAY-2001; 2001WO-US017756.  
XX  
PR 26-MAY-2000; 2000US-0207484P.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
PI Algate PA, Harlocker SL, Jones R;  
XX  
XX WPI; 2002-122075/16.  
XX  
Composition for therapy and diagnosis of ovarian cancer comprising polypeptide of a ovarian tumor polypeptide, polynucleotide encoding polypeptide, antibody specific to polypeptide or T cell expressing polypeptide.

Claim 1; SEQ ID NO 9430; 489pp; English.  
The present invention describes a composition (I) comprising: carriers and immunostimulants; and a polypeptide (II) of a ovarian tumour polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1) from the 10912 nucleotide sequences as given in ABL77023 to ABL87934, or antigen presenting cells that express (II). (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to (S1) can be used for detecting ovarian cancer in a patient's biological sample preferably serum or ovarian tissue. The method comprises contacting a biological sample from a patient with (IV), detecting the amount of polynucleotide hybridising to (IV) and comparing the amount to a predetermined cutoff value and thereby detecting ovarian cancer in the patient, where the amount of polynucleotide hybridising to (IV) is detected preferably by polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is useful for stimulating and/or expanding T cells specific for an ovarian tumour protein comprising contacting T cells with (III) or (II). (III) is useful in design and preparation of ribozyme molecules for inhibiting expression of the tumour polypeptides and proteins in tumour cells; and to isolate a full length gene from a suitable library e.g., a tumour cDNA library using well known techniques

Query Match 2.0%; Score 22; DB 6; Length 141;  
Sequence 141 BP; 58 A; 26 C; 46 G; 11 T; 0 U; 0 Other;  
Best Local Similarity 100.0%; Pred. No. 2;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Best Local Similarity 100.0%; Pred. No. 2;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 862 AGCCTAAAAAGAAAAA 883  
DB 98 AGCCTAAAAAGAAAAA 119

RESULT 3  
ABL86187  
ID ABL86187 standard; cDNA; 144 BP.  
XX  
AC ABL86187;  
XX  
17-MAY-2002 (first entry)  
DE Human ovarian cancer related cDNA clone SEQ ID NO:9165.  
XX  
XX Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.  
XX  
XX Homo sapiens.  
XX  
PN W0200192581-A2.  
XX  
PD 06-DEC-2001.  
XX  
XX 29-MAY-2001; 2001WO-US017756.  
XX  
PR 26-MAY-2000; 2000US-0207484P.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
PI Algate PA, Harlocker SL, Jones R;  
XX  
XX WPI; 2002-122075/16.

Composition for therapy and diagnosis of ovarian cancer comprising polypeptide of a ovarian tumor polypeptide, polynucleotide encoding polypeptide, antibody specific to polypeptide or T cell expressing polypeptide.  
Claim 1; SEQ ID NO 9165; 489pp; English.  
The present invention describes a composition (I) comprising: carriers and immunostimulants; and a polypeptide (II) of a ovarian tumour polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1) from the 10912 nucleotide sequences as given in ABL77023 to ABL87934, or antigen presenting cells that express (II). (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to (S1) can be used for detecting ovarian cancer in a patient's biological sample preferably serum or ovarian tissue. The method comprises contacting a biological sample from a patient with (IV), detecting the amount of polynucleotide hybridising to (IV) and comparing the amount to a predetermined cutoff value and thereby detecting ovarian cancer in the patient, where the amount of polynucleotide hybridising to (IV) is detected preferably by polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is useful for stimulating and/or expanding T cells specific for an ovarian tumour protein comprising contacting T cells with (III) or (II). (III) is useful in design and preparation of ribozyme molecules for inhibiting expression of the tumour polypeptides and proteins in tumour cells; and to isolate a full length gene from a suitable library e.g., a tumour cDNA library using well known techniques

Query Match 2.0%; Score 22; DB 6; Length 144;  
Best Local Similarity 100.0%; Pred. No. 2;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 862 AGCCTAAAAAGAAAAA 883  
DB 99 AGCCTAAAAAGAAAAA 120



KW Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.  
XX Homo sapiens.  
XX WO200170979-A2.  
XX 27-SEP-2001.  
XX 21-MAR-2001; 2001WO-US009126.  
XX 21-MAR-2000; 2000US-0191031P.  
PR 25-MAY-2000; 2000US-0207124P.  
PR 15-JUN-2000; 2000US-0211940P.  
PR 07-JUL-2000; 2000US-0216820P.  
PR 25-JUL-2000; 2000US-0220661P.  
PR 21-DEC-2000; 2000US-0257672P.  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX Lee J, Lillie J;  
PI WPI; 2001-611502/70.  
DR Novel isolated nucleic acid molecules (markers) overexpressed in ovarian  
PT cancer cells as compared to their normal non-cancerous ovarian cells are  
PT used to characterize stage, grade, histological type of ovarian cancer.  
XX Disclosure; SEQ ID NO 20330; 106pp; English.  
XX The invention relates to nucleic acid markers which are overexpressed in  
CC ovarian cancer cells as compared to their expression in normal (i.e. non-  
CC cancerous) ovarian cells. The invention also relates to polypeptides  
CC encoded by the markers, antibodies that selectively bind to the  
CC polypeptides, a method of inhibiting ovarian cancer in a patient at risk  
CC of developing ovarian cancer involving inhibiting expression of a gene  
CC corresponding to a marker of the invention and a method of treating a  
CC patient afflicted with ovarian cancer comprising providing to cells of  
CC the patient an antisense oligonucleotide complementary to a marker of the  
CC invention. The markers are useful for assessing if a patient is afflicted  
CC with ovarian cancer, which involves comparing the level of expression of  
CC a marker in a patient sample and a normal level of expression of the  
CC marker in a control non-ovarian cancer sample. A difference between the  
CC expression levels indicates ovarian cancer. The level of expression of a  
CC marker corresponds to a secreted protein or to a transcribed  
CC polynucleotide or its portion. The level of expression of the marker is  
CC assessed by detecting the presence in the sample, a protein or protein  
CC fragment corresponding to the marker. The presence of protein or protein  
CC fragment is detected using an antibody that specifically binds with the  
CC protein or protein fragment. Alternatively, the level of expression of  
CC the marker is assessed by detecting the presence of a transcribed  
CC polynucleotide which anneals with the marker or anneals with a portion of  
CC the polynucleotide comprising the marker, under stringent conditions. The  
CC marker is also used for monitoring the progression of ovarian cancer in a  
CC patient which involves detecting expression of the marker in a patient  
CC sample at a first point in time, repeating the method at a subsequent  
CC time and comparing the level of expression. The method is carried out  
CC using an ovarian tissue sample. A composition comprising a marker,  
CC polypeptide or antibody of the invention is used to treat ovarian cancer.  
CC This sequence represents a human ovarian cancer DNA marker of the  
XX invention.  
XX Sequence 1074 BP; 194 A; 278 C; 290 G; 300 T; 0 U; 12 Other;  
SQ Query Match 2.0%; Score 22; DB 5; Length 1074;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 867 AAAAAAAAAAAAAAAGTCTT 888  
Db 427 AAAAAAAAAAAAAAAGTCTT 406  
RESULT 7  
Expressed sequence tag; secreted protein; haematopoiesis regulator;  
tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;  
chemotaxis; chemokines; haemostasis; gene therapy; thrombolysis;  
receptor; ligand; anti-inflammatory; tumour inhibitor; ds.  
XX

ABL10390/c  
ID ABL10390 standard; cDNA; 24935 BP.  
XX ABL10390;  
AC ABL10390;  
XX 26-MAR-2002 (first entry)  
DT 26-MAR-2002 (first entry)  
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 25652.  
DE Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX Drosophila melanogaster.  
OS WO200171042-A2.  
XX 27-SEP-2001.  
PD 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US009231.  
PF 23-MAR-2000; 2000US-0191637P.  
XX 11-JUL-2000; 2000US-00614150.  
PR (PEKE ) PE CORP NY.  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI WPI; 2001-656860/75.  
DR P-PSDB; ABB66287.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX Claim 1; SEQ ID NO 25652; 21pp + Sequence Listing; English.  
PS The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX Sequence 24935 BP; 7255 A; 5321 C; 5280 G; 7079 T; 0 U; 0 Other;  
SQ Query Match 2.0%; Score 22; DB 4; Length 24935;  
Best Local Similarity 100.0%; Pred. No. 2.4;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 872 AAGAAAAAAACTGTTCATC 893  
Db 12697 AAGAAAAAAACTGTTCATC 12676  
RESULT 8  
AAV86132/c  
ID AAV86132 standard; cDNA; 377 BP.  
XX AAV86132;  
AC AAV86132;  
XX 27-APR-1999 (first entry)  
DT 27-APR-1999 (first entry)  
XX EST clone H165.  
DE Expressed sequence tag; secreted protein; haematopoiesis regulator;  
XX tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;  
KW chemotaxis; chemokines; haemostasis; gene therapy; thrombolysis;  
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.  
XX

```

OS Homo sapiens.
XX WO9845435-A2.
XX
XX 15-OCT-1998.
XX
XX 10-APR-1998; 98WO-US006954.
XX
XX 10-APR-1997; 97US-00835913.
XX
XX (GEM ) GENETICS INST INC.
XX
XX Jacobs K, McCoy JM, Lavallie ER, Racie LA, Merberg D, Treacy M;
XX Spaulding V, Agostino MJ;
XX WPI; 1999-070076/06.
XX
XX New polynucleotides encoding human secreted proteins - derived from e.g.
XX human blood, kidney, foetal lung, placenta, testes, brain, ovary,
XX pituitary, retina and colon cDNA libraries.
XX
XX Claim 1; Page 132; 633pp; English.
XX
XX This sequence represents an expressed sequence tag (EST), and is a
XX polynucleotide of the invention. The polynucleotides of the invention are
XX all secreted EST sequences isolated from a variety of human tissue
XX sources. The EST sequences and proteins encoded by them are predicted to
XX have useful biological activities which would make them suitable for
XX treating, preventing or ameliorating medical conditions in humans and
XX animals, although no supporting data is given. Suggested activities
XX include nutritional activity, immune stimulating or suppressing activity,
XX haematopoiesis regulating activity, tissue growth activity,
XX activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, receptor/ligand activity, anti-inflammatory
XX activity, cadherin/tumour invasion suppressor activity, tumour inhibition
XX activity. The EST sequences are also stated to be useful for gene therapy
XX
XX Sequence 377 BP; 89 A; 88 C; 76 G; 124 T; 0 U; 0 Other;

Query Match 1.9%; Score 21; DB 2; Length 377;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 863 GCCTAAAAAAGAAAAA 883
Db 43 GCCTAAAAAAGAAAAA 23

RESULT 9
AAC24900
ID AAC24900 standard; cDNA; 403 BP.
XX
XX AAC24900;
XX
XX 06-OCT-2000 (first entry)
XX
XX Human secreted protein 5' EST, SEQ ID NO: 28975.
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping; ss.
XX
XX Homo sapiens.
XX
XX EP1033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-00200610.
XX
XX 26-FEB-1999; 99US-0122487P.
XX
XX (GEST ) GENSET.
XX

PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
XX Claim 1; SEQ ID NO 28975; 71pp + Sequence Listing; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
XX ESTs are derived from mRNAs with intact 5' ends and can therefore be used
XX to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
XX diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors
XX
XX Sequence 403 BP; 141 A; 68 C; 60 G; 128 T; 0 U; 6 Other;

Query Match 1.9%; Score 21; DB 3; Length 403;
Best Local Similarity 100.0%; Pred. No. 6.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 863 GCCTAAAAAAGAAAAA 883
Db 163 GCCTAAAAAAGAAAAA 183

RESULT 10
AAH71551
ID AAH71551 standard; cDNA; 621 BP.
XX
XX AAH71551;
XX
XX 19-SEP-2001 (first entry)
XX
XX Human cervical cancer marker nucleic acid 2825.
XX
XX Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX WO200142467-A2.
XX
XX 14-JUN-2001.
XX
XX 08-DEC-2000; 2000WO-US033312.
XX
XX 08-DEC-1999; 99US-0169681P.
XX 21-DEC-1999; 99US-0171350P.
XX 14-MAR-2000; 2000US-0189315P.
XX 12-MAY-2000; 2000US-0203791P.
XX 09-JUN-2000; 2000US-0210600P.
XX 21-JUL-2000; 2000US-0220114P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Schlegel R, Deeds J, Berger A, Zhao X;
XX WPI; 2001-375006/39.
XX
XX New isolated nucleic acid for diagnosing and treating cervical cancer and
XX for assessing and detecting compounds for treating the cancer.
XX
XX Claim 1; Page 579; 1051pp; English.

```

```
XX The invention relates to novel genes (AAH68727-AAH73383) associated with
CC cervical cancer with cytostatic activity. The nucleic acids and encoded
CC polypeptides are useful: to assess if a patient is afflicted with
CC cervical cancer or has a pre-malignant condition; to monitor the
CC progression of cervical cancer or a premalignant condition in a patient;
CC and to select and/or assess the efficacy of a compound or therapy for
CC inhibiting cervical cancer in a patient. The nucleic acids may also be
CC useful for gene therapy
XX
SQ Sequence 621 BP; 273 A; 63 C; 47 G; 238 T; 0 U; 0 Other;
Query Match 1.9%; Score 21; DB 4; Length 621;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 864 CCTAAGAAAAAGAAAAAAGAAC 884
DB 601 CCTAAGAAAAAGAAAAAAGAAC 621
RESULT 11
AAF16195/c
ID AAF16195 standard; cDNA; 1097 BP.
XX
AC AAF16195;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human prostate cancer antigen nucleotide sequence SEQ ID NO:630.
XX
KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
KW vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease; ss.
XX
OS Homo sapiens.
XX
PN WO200055174-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US005988.
XX
PR 12-MAR-1999; 99US-0124270P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-587513/55.
DR P-PSDB; AAB56992.
XX
PT Prostate cancer associated gene sequences, referred to as prostate cancer
PT antigens, useful for treatment, prevention, and diagnosis of disorders
PT such as prostate cancer.
XX
PS Claim 1; Page 1072-1073; 2338pp; English.
XX
CC AAP15566 to AAP16505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytostatic,
CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,
CC nephrotropic, antiinfective, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
```

---

```
CC disorders, wounds, and infectious diseases. AAP16506 to AAP16514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention
XX
SQ Sequence 1097 BP; 301 A; 204 C; 201 G; 389 T; 0 U; 2 Other;
Query Match 1.9%; Score 21; DB 3; Length 1097;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 863 GCTTAAAAAGAAAAAAGAAA 883
DB 162 GCTTAAAAAGAAAAAAGAAA 142
RESULT 12
ADA72818/c
ID ADA72818 standard; DNA; 2000 BP.
XX
AC ADA72818;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 6143.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX
OS Oryza sativa.
XX
PN WO2003000898-A1.
XX
PD 03-JAN-2003.
XX
PF 22-JUN-2001; 2001WO-IB001105.
XX
PR 22-JUN-2001; 2001WO-IB001105.
XX
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX
DR WPI; 2003-175290/17.
XX
PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
PS Claim 27; SEQ ID NO 6143; 899pp; English.
XX
CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 2000 BP; 632 A; 370 C; 326 G; 672 T; 0 U; 0 Other;
Query Match 1.9%; Score 21; DB 8; Length 2000;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 867 AAAAAAGAAAAAAACTGT 887
DB 1385 AAAAAAGAAAAAAACTGT 1365
```

```
RESULT 13
AAH72860/c
ID AAH72860 standard; cDNA; 2170 BP.
XX
XX AAH72880;
XX
XX
XX
XX 19-SEP-2001 (first entry)
XX
XX Human cervical cancer marker nucleic acid 4154.
DE
XX
XX Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200142467-A2.
XX
XX 14-JUN-2001.
XX
XX 08-DEC-2000; 2000WO-US033312.
XX
XX 08-DEC-1999; 99US-0169681P.
XX
XX 21-DEC-1999; 99US-0171350P.
XX
XX 14-MAR-2000; 2000US-0189315P.
XX
XX 12-MAY-2000; 2000US-0203791P.
XX
XX 09-JUN-2000; 2000US-0210600P.
XX
XX 21-JUL-2000; 2000US-0220114P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Schlegel R, Deeds J, Berger A, Zhao X;
XX
XX WPI; 2001-375006/39.
XX
XX New isolated nucleic acid for diagnosing and treating cervical cancer and
XX for assessing and detecting compounds for treating the cancer.
XX
XX Claim 1; Page 903; 1051pp; English.
XX
XX The invention relates to novel genes (AAH68727-AAH73383) associated with
XX cervical cancer with cytostatic activity. The nucleic acids and encoded
XX polypeptides are useful: to assess if a patient is afflicted with
XX cervical cancer or has a pre-malignant condition; to monitor the
XX progression of cervical cancer or a premalignant condition in a patient;
XX and to select and/or assess the efficacy of a compound or therapy for
XX inhibiting cervical cancer in a patient. The nucleic acids may also be
XX useful for gene therapy
XX
XX Sequence 2170 BP; 358 A; 689 C; 427 G; 688 T; 0 U; 8 Other;
SQ
Query Match 1.9%; Score 21; DB 4; Length 2170;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 866 TAAAAAAGAAAAAAGAACTG 886
DB 96 TAAAAAAGAAAAAAGAACTG 76

RESULT 14
ABV22589/c
ID ABV22589 standard; cDNA; 2170 BP.
XX
XX ABV22589;
XX
XX 13-SEP-2002 (first entry)
XX
XX Human prostate expression marker cDNA 22580.
XX
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX pharmacogenomic marker; gene; ss.
XX
XX Homo sapiens.
OS
```

```
XX WO200160860-A2.
XX
XX 23-AUG-2001.
XX
XX 20-FEB-2001; 2001WO-US005171.
XX
XX 17-FEB-2000; 2000US-0183319P.
XX
XX 16-MAR-2000; 2000US-0189862P.
XX
XX 25-MAY-2000; 2000US-0207454P.
XX
XX 09-JUN-2000; 2000US-0211314P.
XX
XX 18-JUL-2000; 2000US-0219007P.
XX
XX 13-DEC-2000; 2000US-0255281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Schlegel R, Endege WO, Monahan JE;
XX
XX WPI; 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
XX prostate cells and correlating with presence of prostate cancer, useful
XX for detecting presence of prostate cancer, stage of prostate cancer.
XX
XX Claim 1; Page 3950; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
XX a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX specification or its complement. (I) is useful for: (a) assessing whether
XX a patient is afflicted with prostate cancer; (b) monitoring the efficacy
XX progression of prostate cancer in a patient; (c) assessing the efficacy
XX of a test compound to inhibit prostate cancer in a patient; (d) assessing
XX the efficacy of a therapy for inhibiting prostate cancer in a patient;
XX (e) selecting a composition for inhibiting prostate cancer in a patient;
XX (f) assessing the prostate cell carcinogenic potential of a compound; (g)
XX determining whether prostate cancer has metastasized in a patient; (h)
XX assessing the aggressiveness or indolence of prostate cancer in a patient
XX ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
XX Sequence 2170 BP; 358 A; 688 C; 428 G; 688 T; 0 U; 8 Other;
SQ
Query Match 1.9%; Score 21; DB 5; Length 2170;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 866 TAAAAAAGAAAAAAGAACTG 886
DB 96 TAAAAAAGAAAAAAGAACTG 76

RESULT 15
ABV25695/c
ID ABV25695 standard; cDNA; 2170 BP.
XX
XX ABV25695;
XX
XX 16-SEP-2002 (first entry)
XX
XX Human prostate expression marker cDNA 25686.
XX
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX pharmacogenomic marker; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200160860-A2.
XX
XX 23-AUG-2001.
XX
XX 20-FEB-2001; 2001WO-US005171.
XX
XX 17-FEB-2000; 2000US-0183319P.
XX
XX 16-MAR-2000; 2000US-0189862P.
XX
```







Result No.	Score	Query	Length	DB	ID	Description	
C 1	22	2.0	249	4	US-09-248-796A-9416	Sequence 9416, Ap	
C 2	22	2.0	601	4	US-09-949-016-119414	Sequence 119414,	
C 3	22	2.0	601	4	US-09-949-016-119415	Sequence 119415,	
C 4	22	2.0	51403	4	US-09-949-016-15057	Sequence 15057, A	
C 5	22	2.0	250352	4	US-09-949-016-14724	Sequence 14724, A	
C 6	21	1.9	403	4	US-09-513-999C-28975	Sequence 28975, A	
C 7	21	1.9	601	4	US-09-949-016-120940	Sequence 120940,	
C 8	21	1.9	601	4	US-09-949-016-169986	Sequence 169986,	
C 9	21	1.9	33908	4	US-09-949-016-15104	Sequence 15104, A	
C 10	21	1.9	114139	4	US-09-949-016-18536	Sequence 18536, A	
C 11	20	1.8	601	4	US-09-949-016-80010	Sequence 80010, A	
C 12	20	1.8	601	4	US-09-949-016-80011	Sequence 80011, A	
C 13	20	1.8	601	4	US-09-949-016-80012	Sequence 80012, A	
C 14	20	1.8	601	4	US-09-949-016-136288	Sequence 136288,	
C 15	20	1.8	601	4	US-09-949-016-148215	Sequence 148215,	
C 16	20	1.8	601	4	US-09-949-016-148216	Sequence 148216,	
C 17	20	1.8	667	4	US-09-771-035A-18	Sequence 18, Appl	
C 18	20	1.8	1725	4	US-09-949-016-3377	Sequence 3377, Ap	
C 19	20	1.8	1890	4	US-09-949-016-4824	Sequence 4824, Ap	
C 20	20	1.8	8625	4	US-09-949-016-16566	Sequence 16566, A	
C 21	20	1.8	13985	4	US-09-949-016-15640	Sequence 15640, A	
C 22	20	1.8	15632	4	US-09-949-016-15119	Sequence 15119, A	
C 23	20	1.8	22471	4	US-09-949-016-15902	Sequence 15902, A	
C 24	20	1.8	91772	4	US-09-949-016-15568	Sequence 15568, A	
C 25	20	1.8	154023	4	US-09-949-016-17057	Sequence 17057, A	
C 26	20	1.8	171700	4	US-09-949-016-12276	Sequence 12276, A	
C 27	20	1.8	171701	4	US-09-949-016-15835	Sequence 15835, A	

```
; SEQ ID NO 119414
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-119414

Query Match          2.0%; Score 22; DB 4; Length 601;
Best Local Similarity 100.0%; Pred. No. 0.77;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 868 AAAAAAGAAAAAACTGTTTC 889
    |||||
Db 378 AAAAAAGAAAAAACTGTTTC 357

RESULT 3
US-09-949-016-119415/c
; Sequence 119415, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 119415
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-119415

Query Match          2.0%; Score 22; DB 4; Length 601;
Best Local Similarity 100.0%; Pred. No. 0.77;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 868 AAAAAAGAAAAAACTGTTTC 889
    |||||
Db 60 AAAAAAGAAAAAACTGTTTC 39

RESULT 4
US-09-949-016-15057/c
; Sequence 15057, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15057
; LENGTH: 51403
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
```

```
; NAME/KEY: misc feature
; LOCATION: (1)-(51403)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15057

Query Match          2.0%; Score 22; DB 4; Length 51403;
Best Local Similarity 100.0%; Pred. No. 0.82;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 868 AAAAAAGAAAAAACTGTTTC 889
    |||||
Db 47891 AAAAAAGAAAAAACTGTTTC 47870

RESULT 5
US-09-949-016-14724/c
; Sequence 14724, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14724
; LENGTH: 250352
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(250352)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14724

Query Match          2.0%; Score 22; DB 4; Length 250352;
Best Local Similarity 100.0%; Pred. No. 0.83;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 865 CTAAGAAAAAGAAAAAACTG 886
    |||||
Db 228042 CTAAGAAAAAGAAAAAACTG 228021

RESULT 6
US-09-513-999C-28975
; Sequence 28975, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; PATENT NO. 6783961
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 28975
; LENGTH: 403
; TYPE: DNA
; ORGANISM: Homo sapiens
```



```
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16536
; LENGTH: 114139
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc feature
; LOCATION: (1)...(114139)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16536

Query Match      1.9%; Score 21; DB 4; Length 114139;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      868 AAAAAAGAAAAAACTGTT 888
Db      57687 AAAAAAGAAAAAACTGTT 57707

RESULT 11
US-09-949-016-80010/c
; Sequence 80010, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 80010
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-80010

Query Match      1.8%; Score 20; DB 4; Length 601;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      864 CCTAAAAAGAAAAAAA 883
Db      368 CCTAAAAAGAAAAAAA 349

RESULT 12
US-09-949-016-80011/c
; Sequence 80011, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 80011
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-80011

Query Match      1.8%; Score 20; DB 4; Length 601;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      864 CCTAAAAAGAAAAAAA 883
Db      220 CCTAAAAAGAAAAAAA 201

RESULT 13
US-09-949-016-80012/c
; Sequence 80012, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 80012
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-80012

Query Match      1.8%; Score 20; DB 4; Length 601;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      864 CCTAAAAAGAAAAAAA 883
Db      220 CCTAAAAAGAAAAAAA 201

RESULT 14
US-09-949-016-136288/c
; Sequence 136288, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
```

; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 136288  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-136288

Query Match 1.8%; Score 20; DB 4; Length 601;  
Best Local Similarity 100.0%; Pred. No. 7.9;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 875 AAAAAAAAACTGTTCAATCA 894  
Db 309 AAAAAAAAACTGTTCAATCA 290

RESULT 15

US-09-949-016-148215/c  
; Sequence 148215, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 148215  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-148215

Query Match 1.8%; Score 20; DB 4; Length 601;  
Best Local Similarity 100.0%; Pred. No. 7.9;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 416 GGAGGAGAAAGTAGGA 435  
Db 335 GGAGGAGAAAGTAGGA 316

Search completed: August 30, 2005, 19:38:33  
Job time : 231 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 30, 2005, 16:45:56 ; Search time 859 Seconds  
(without alignments)  
8462.599 Million cell updates/sec

Title: US-09-979-549-2

Perfect score: 1111

Sequence: 1 gaagtcaagaagccgttcag.....ccgagaagagagagagag 1111

Scoring table:

OLIGO NUC

Gapop 60.0 , Gapext 60.0

Searched: 7331713 seqs, 3271544945 residues

Word size : 0

Total number of hits satisfying chosen parameters: 14663426

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*
- 17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq.\*
- 18: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq.\*
- 19: /cgn2\_6/ptodata/1/pubpna/US10G\_PUBCOMB.seq.\*
- 20: /cgn2\_6/ptodata/1/pubpna/US10H\_PUBCOMB.seq.\*
- 21: /cgn2\_6/ptodata/1/pubpna/US10I\_PUBCOMB.seq.\*
- 22: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 23: /cgn2\_6/ptodata/1/pubpna/US11A\_PUBCOMB.seq.\*
- 24: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq.\*
- 25: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 26: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	3.8	629	20	US-10-425-115-15660
2	22	2.0	141	9	Sequence 15660, A
3	22	2.0	144	9	Sequence 9430, Ap
4	22	2.0	154	9	Sequence 9165, Ap
5	22	2.0	213	9	Sequence 9299, Ap
6	22	2.0	1074	10	Sequence 9926, Ap
7	22	2.0	394468	21	Sequence 20330, A
					Sequence 17952, A

8	21	1.9	261	17	US-10-242-535A-46223	Sequence 46223, A
9	21	1.9	261	18	US-10-085-783A-46223	Sequence 46223, A
10	21	1.9	542	20	US-10-425-115-101740	Sequence 101740, A
11	21	1.9	614	19	US-10-437-963-3058	Sequence 3058, Ap
12	21	1.9	828	13	US-10-027-632-169396	Sequence 169396, A
13	21	1.9	828	17	US-10-027-632-169396	Sequence 169396, A
14	21	1.9	1097	9	US-09-925-300-630	Sequence 630, App
15	21	1.9	1420	19	US-10-437-963-85377	Sequence 85377, A
16	21	1.9	2170	20	US-10-357-930-22582	Sequence 22582, A
17	21	1.9	2170	20	US-10-357-930-24738	Sequence 24738, A
18	21	1.9	2170	20	US-10-357-930-25684	Sequence 25684, A
19	21	1.9	2170	20	US-10-357-930-28422	Sequence 28422, A
20	21	1.9	3191	10	US-09-374-046A-59	Sequence 59, Appl
21	21	1.9	3191	18	US-10-616-263-59	Sequence 59, Appl
22	21	1.9	3618	18	US-10-380-374-21	Sequence 21, Appl
23	21	1.9	4740	20	US-10-357-930-21715	Sequence 21715, A
24	21	1.9	4740	20	US-10-357-930-27560	Sequence 27560, A
25	21	1.9	10351	9	US-09-874-470-5	Sequence 5, Appli
26	21	1.9	80815	19	US-10-322-281-486	Sequence 486, App
27	20	1.8	162	17	US-10-131-827-8661	Sequence 8661, Ap
28	20	1.8	369	13	US-10-027-632-68718	Sequence 68718, A
29	20	1.8	369	17	US-10-027-632-68718	Sequence 68718, A
30	20	1.8	371	10	US-09-960-706-561	Sequence 561, App
31	20	1.8	371	10	US-09-873-319-346	Sequence 346, App
32	20	1.8	375	20	US-10-357-930-48707	Sequence 48707, A
33	20	1.8	422	11	US-09-969-034-2636	Sequence 2636, Ap
34	20	1.8	462	10	US-09-918-995-9917	Sequence 9917, Ap
35	20	1.8	555	9	US-09-764-869-2246	Sequence 2246, Ap
36	20	1.8	555	14	US-10-091-504-2246	Sequence 2246, Ap
37	20	1.8	555	17	US-10-227-577-2246	Sequence 2246, Ap
38	20	1.8	571	13	US-10-027-632-47881	Sequence 47881, A
39	20	1.8	571	13	US-10-027-632-295267	Sequence 295267, A
40	20	1.8	571	17	US-10-027-632-47881	Sequence 47881, A
41	20	1.8	571	17	US-10-027-632-295267	Sequence 295267, A
42	20	1.8	630	13	US-10-027-632-280598	Sequence 280598, A
43	20	1.8	630	17	US-10-027-632-280598	Sequence 280598, A
44	20	1.8	667	9	US-09-771-035A-18	Sequence 18, Appl
45	20	1.8	667	18	US-10-634-221-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1

US-10-425-115-15660  
; Sequence 15660, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 15660  
; LENGTH: 629  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_114281C.1  
US-10-425-115-15660

Query Match 3.8%; Score 42; DB 20; Length 629;  
Best Local Similarity 100.0%; Pred. No. 5.3e-12;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1033 GAGACAGAGCCCAAGGCAACAAAGTCGCGTGAGAAA 1074

DB 26 GAGACAGAGCCCAAGGCAACAAAGTCGCGTGAGAAA 67

```
RESULT 2
US-09-867-701-9430
; Sequence 9430, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9430
; LENGTH: 141
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9430

Query Match      2.0%; Score 22; DB 9; Length 141;
Best Local Similarity 100.0%; Pred. No. 0.62;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      862 AGCCTAAAAAAGAAAAA 883
Db      98 AGCCTAAAAAAGAAAAA 119

RESULT 3
US-09-867-701-9165
; Sequence 9165, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9165
; LENGTH: 144
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9165

Query Match      2.0%; Score 22; DB 9; Length 144;
Best Local Similarity 100.0%; Pred. No. 0.62;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      862 AGCCTAAAAAAGAAAAA 883
Db      99 AGCCTAAAAAAGAAAAA 120

RESULT 4
US-09-867-701-9299
; Sequence 9299, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
```

```
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9299
; LENGTH: 154
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9299

Query Match      2.0%; Score 22; DB 9; Length 154;
Best Local Similarity 100.0%; Pred. No. 0.62;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      862 AGCCTAAAAAAGAAAAA 883
Db      112 AGCCTAAAAAAGAAAAA 133

RESULT 5
US-09-867-701-9926
; Sequence 9926, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9926
; LENGTH: 213
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9926

Query Match      2.0%; Score 22; DB 9; Length 213;
Best Local Similarity 100.0%; Pred. No. 0.63;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      862 AGCCTAAAAAAGAAAAA 883
Db      163 AGCCTAAAAAAGAAAAA 184

RESULT 6
US-09-814-353-20330/c
; Sequence 20330, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
```



;; PRIOR APPLICATION NUMBER: US 60/257,672  
;; PRIOR FILING DATE: 2000-12-21  
;; NUMBER OF SEQ ID NOS: 22037  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 20330  
;; LENGTH: 1074  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: misc\_feature  
;; LOCATION: 1..2..3..4..5..6..7..8..9..10..968..1074  
;; OTHER INFORMATION: n = A,T,C or G  
US-09-814-353-20330

Query Match 2.0%; Score 22; DB 10; Length 1074;  
Best Local Similarity 100.0%; Pred. No. 0.69;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 867 AAAAAAGAAAAAACTGTT 888  
Db 427 AAAAAAGAAAAAACTGTT 406

RESULT 7  
US-10-741-600-17952/c  
;; Sequence 17952, Application US/10741600  
;; Publication No. US20050026169A1  
;; GENERAL INFORMATION:  
;; APPLICANT: CARGILL, Michele et al.  
;; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
;; FILE REFERENCE: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF  
;; FILE REFERENCE: CL001499  
;; CURRENT APPLICATION NUMBER: US/10/741,600  
;; CURRENT FILING DATE: 2003-12-22  
;; NUMBER OF SEQ ID NOS: 73997  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 17952  
;; LENGTH: 394468  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: misc\_feature  
;; LOCATION: (1)...(394468)  
;; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-  
US-10-741-600-17952

Query Match 2.0%; Score 22; DB 21; Length 394468;  
Best Local Similarity 100.0%; Pred. No. 0.92;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 865 CTAAGAAAAAGAAAAAACTG 886  
Db 53760 CTAAGAAAAAGAAAAAACTG 53739

RESULT 8  
US-10-242-535A-46223  
;; Sequence 46223, Application US/10242535A  
;; Publication No. US20040013663A1  
;; GENERAL INFORMATION:  
;; APPLICANT: ChondroGene Inc.  
;; APPLICANT: Liew, C.C.  
;; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
;; FILE REFERENCE: 4231/2005  
;; CURRENT APPLICATION NUMBER: US/10/242,535A  
;; CURRENT FILING DATE: 2002-09-12  
;; PRIOR APPLICATION NUMBER: US 10/085,783  
;; PRIOR FILING DATE: 2002-02-28  
;; PRIOR APPLICATION NUMBER: US 60/305,340  
;; PRIOR FILING DATE: 2001-07-13  
;; PRIOR APPLICATION NUMBER: US 60/275,017  
;; PRIOR FILING DATE: 2001-03-12  
;; PRIOR APPLICATION NUMBER: US 60/271,955

;; PRIOR FILING DATE: 2001-02-28  
;; NUMBER OF SEQ ID NOS: 58994  
;; SOFTWARE: PatentIn version 3.2  
;; SEQ ID NO 46223  
;; LENGTH: 261  
;; TYPE: DNA  
;; ORGANISM: Human  
US-10-242-535A-46223

Query Match 1.9%; Score 21; DB 17; Length 261;  
Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 863 GCCTAAAAAGAAAAAAA 883  
Db 222 GCCTAAAAAGAAAAAAA 242

RESULT 9  
US-10-085-783A-46223  
;; Sequence 46223, Application US/10085783A  
;; Publication No. US20040037841A1  
;; GENERAL INFORMATION:  
;; APPLICANT: ChondroGene Inc.  
;; APPLICANT: Liew, C.C.  
;; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
;; FILE REFERENCE: 4231/2002  
;; CURRENT APPLICATION NUMBER: US/10/085,783A  
;; CURRENT FILING DATE: 2002-02-28  
;; PRIOR APPLICATION NUMBER: US 60/305,340  
;; PRIOR FILING DATE: 2001-07-13  
;; PRIOR APPLICATION NUMBER: US 60/275,017  
;; PRIOR FILING DATE: 2001-03-12  
;; PRIOR APPLICATION NUMBER: US 60/271,955  
;; PRIOR FILING DATE: 2001-02-28  
;; NUMBER OF SEQ ID NOS: 58994  
;; SOFTWARE: PatentIn version 3.2  
;; SEQ ID NO 46223  
;; LENGTH: 261  
;; TYPE: DNA  
;; ORGANISM: Human  
US-10-085-783A-46223

Query Match 1.9%; Score 21; DB 18; Length 261;  
Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 863 GCCTAAAAAGAAAAAAA 883  
Db 222 GCCTAAAAAGAAAAAAA 242

RESULT 10  
US-10-425-115-101740/c  
;; Sequence 101740, Application US/10425115  
;; Publication No. US20040214272A1  
;; GENERAL INFORMATION:  
;; APPLICANT: La Rosa, Thomas J.  
;; APPLICANT: Kovalic, David K.  
;; APPLICANT: Zhou, Yihua  
;; APPLICANT: Cao, Yongwei  
;; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
;; FILE REFERENCE: 38-21(53222)B  
;; CURRENT APPLICATION NUMBER: US/10/425,115  
;; CURRENT FILING DATE: 2003-04-28  
;; NUMBER OF SEQ ID NOS: 369326  
;; SEQ ID NO 101740  
;; LENGTH: 542  
;; TYPE: DNA  
;; ORGANISM: Zea mays  
;; FEATURE:  
;; OTHER INFORMATION: Clone ID: MMT4577\_24297C.1

```
US-10-425-115-101740
Query Match          1.9%; Score 21; DB 20; Length 542;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      863 GCCTAAAAAAGAAAAA 883
Db      50 GCCTAAAAAAGAAAAA 30

RESULT 11
US-10-437-963-3058/c
; Sequence 3058, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 3058
; LENGTH: 614
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_102776C.1
US-10-437-963-3058

Query Match          1.9%; Score 21; DB 19; Length 614;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1091 GCCGAGAAGAGAGAGAGAG 1111
Db      108 GCCGAGAAGAGAGAGAGAG 88

RESULT 12
US-10-632-169396
; Sequence 169396, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 169396
; LENGTH: 828
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-169396

Query Match          1.9%; Score 21; DB 17; Length 828;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      867 AAAAAAGAAAAAACTGT 887
Db      118 AAAAAAGAAAAAACTGT 138

RESULT 13
US-10-027-632-169396
; Sequence 169396, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 169396
; LENGTH: 828
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-169396

Query Match          1.9%; Score 21; DB 17; Length 828;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      867 AAAAAAGAAAAAACTGT 887
Db      118 AAAAAAGAAAAAACTGT 138

RESULT 14
US-09-925-300-630/c
; Sequence 630, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
```

```
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 169396
; LENGTH: 828
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-169396

Query Match          1.9%; Score 21; DB 13; Length 828;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      867 AAAAAAGAAAAAACTGT 887
Db      118 AAAAAAGAAAAAACTGT 138

RESULT 13
US-10-027-632-169396
; Sequence 169396, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 169396
; LENGTH: 828
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-169396

Query Match          1.9%; Score 21; DB 17; Length 828;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      867 AAAAAAGAAAAAACTGT 887
Db      118 AAAAAAGAAAAAACTGT 138

RESULT 14
US-09-925-300-630/c
; Sequence 630, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
```

; NUMBER OF SEQ ID NOS: 1890  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 630  
; LENGTH: 1097  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (13)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-925-300-630

Query Match 1.9%; Score 21; DB 9; Length 1097;  
Best Local Similarity 100.0%; Pred. No. 2.5;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 863 GCCTAAGAAAAAGAAAAAA 883  
DB 162 GCCTAAGAAAAAGAAAAAA 142

## RESULT 15

US-10-437-963-85377  
; Sequence 85377, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 85377  
; LENGTH: 1420  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(1420)  
; OTHER INFORMATION: unsure at all n locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_84521C.1  
US-10-437-963-85377

Query Match 1.9%; Score 21; DB 19; Length 1420;  
Best Local Similarity 100.0%; Pred. No. 2.5;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 870 AAAAGAAAAAACTGTTCA 890  
DB 1363 AAAAGAAAAAACTGTTCA 1383

Search completed: August 30, 2005, 19:52:58  
Job time : 860 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 30, 2005, 16:27:35 ; Search time 4253 Seconds  
(without alignments)  
9943.429 Million cell updates/sec

Title: US-09-979-549-2

Perfect score: 1111

Sequence: 1 gaagtcaagcgcttcag.....ccgagaagagagagagag 1111

Scoring table: OLIGO NUC  
Gapop\_60.0 , Gapext 60.0

Searched: 34239544 seqs, 19032134700 residues

Word size : 0

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:\*

- 1: gb\_est1:\*
- 2: gb\_est2:\*
- 3: gb\_hic:\*
- 4: gb\_est3:\*
- 5: gb\_est4:\*
- 6: gb\_est5:\*
- 7: gb\_est6:\*
- 8: gb\_ges1:\*
- 9: gb\_ges2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	99	8.9	509	6	CD867573
2	99	8.9	570	6	CD867174
3	92	8.3	594	6	CA593541
4	90	8.3	631	6	CD938039
5	90	8.1	624	6	CD865240
6	88	7.9	430	6	CA712930
7	66	5.9	643	6	CD884411
8	64	5.8	642	7	CF132917
9	64	5.8	712	6	CD894217
10	63	5.7	556	6	CA701748
11	63	5.7	637	6	CD904051
12	62	5.6	422	2	BE419251
13	62	5.6	578	4	BJ257084
14	62	5.6	579	4	BJ221033
15	60	5.4	597	4	BJ253061
16	58	5.2	533	4	BJ239800
17	58	5.2	585	4	BJ233935
18	56	5.0	543	6	CA497278
19	56	5.0	1061	7	CK213031
20	55	5.0	593	4	BJ296612
21	55	5.0	605	4	BJ290275
22	54	4.9	582	4	BJ262654
23	54	4.9	875	7	CK159678
24	53	4.8	839	7	CK158306

C	25	53	4.8	889	7	CK158740	CK158740 FGAS04003
	26	52	4.7	418	6	CA702286	CA702286 wdk1c.pk0
	27	47	4.2	557	4	BJ223826	BJ223826 BJ223826
	28	47	4.2	628	5	BQ805515	BQ805515 WHE3567_H
	29	44	4.0	353	2	BE406240	BE406240 WHE0409_D
C	30	43	3.9	563	4	BJ273554	BJ273554 BJ273554
	31	43	3.9	565	5	BU101004	BU101004 WHE3360_F
	32	43	3.9	567	6	CA717478	CA717478 wdk4c.pk0
C	34	43	3.9	630	4	BJ320546	BJ320546 BJ320546
	35	43	3.9	630	6	CD930486	CD930486 GR45.111H
	36	43	3.9	644	6	CD874126	CD874126 AZO3.101H
C	37	43	3.9	1141	7	CK211149	CK211149 FGAS02298
	38	41	3.7	459	1	AL826362	AL826362 AL826362
	39	40	3.6	156	2	BE419352	BE419352 WMS01.ASR
	40	40	3.6	608	4	BJ223851	BJ223851 BJ223851
	41	40	3.6	633	4	BJ258073	BJ258073 BJ258073
	42	34	3.1	577	4	BJ214045	BJ214045 BJ214045
C	43	34	3.1	887	7	CK158744	CK158744 FGAS04003
	44	32	2.9	137	4	BJ223280	BJ223280 BJ223280
	45	32	2.9	160	4	BJ290875	BJ290875 BJ290875

ALIGNMENTS

RESULT 1  
CD867573 509 bp mRNA linear EST 11-JUL-2003  
LOCUS AZO2.106K05F001108 AZO2 Triticum aestivum cDNA clone AZO2106K05,  
DEFINITION mRNA sequence.  
ACCESSION CD867573  
VERSION CD867573.1 GI:32551389  
KEYWORDS EST.  
SOURCE Triticum aestivum (bread wheat)  
ORGANISM Triticum aestivum  
REFERENCE 1 (bases 1 to 509)  
AUTHORS Genoplante.  
TITLE Genoplante, a major partnership french program in plant genomics  
JOURNAL Unpublished (2003)  
COMMENT Contact: Genoplante  
Genoplante  
93, rue Henri Rochefort 91025 EVRY CEDEX France  
Tel: 33 1 69 47 54 00  
Fax: 33 1 69 47 54 10  
This sequence has been generated in the framework of the french  
plant genomics programme 'Genoplante' (<http://www.genoplante.com>  
and <http://genoplante-info.infobiogen.fr>).

FEATURES

Location/Qualifiers  
1..509  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="recital"  
/db\_xref="taxon:4565"  
/clone="AZO2106K05"  
/tissue\_type="root"  
/clone\_lib="AZO2"

ORIGIN

Query Match 8.9%; Score 99; DB 6; Length 509;  
Best Local Similarity 100.0%; Pred. No. 1.4e-41;  
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1013 GGGAGCGCGCGCGGATTGGAGACAGACCCCAAGGCAACAAAGTCCGCGTGAGA 1072  
Dbb 1 GGGAGCGCGCGCGGATTGGAGACAGACCCCAAGGCAACAAAGTCCGCGTGAGA 60  
QY 1073 AATCAACAGCGGTGCTTCCGAGAGAGAGAGAGAG 1111  
Dbb 61 AATCAACAGCGGTGCTTCCGAGAGAGAGAGAGAG 99

```
RESULT 2
CD867174
LOCUS
DEFINITION AZ02.105301F001124 AZ02 Triticum aestivum cDNA clone AZ02105J01,
mRNA sequence.
ACCESSION CD867174
VERSION CD867174.1 GI:32550990
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
1 (bases 1 to 570)
REFERENCE
AUTHORS Genoplante.
TITLE Genoplante, a major partnership french program in plant genomics
JOURNAL Unpublished (2003)
COMMENT Contact: Genoplante
Genoplante
93, rue Henri Rochefort 91025 EVRY CEDEX France
Tel: 33 1 69 47 54 00
Fax: 33 1 69 47 54 10
This sequence has been generated in the framework of the french
plant genomics programme 'Genoplante' (http://www.genoplante.com
and http://genoplante-info.infobiogen.fr).
location/Qualifiers
source
1..570
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="recital"
/db_xref="taxon:4565"
/clone="AZ02105J01"
/tissue_type="root"
/clone_lib="AZ02"

ORIGIN
Query Match 8.9%; Score 99; DB 6; Length 570;
Best Local Similarity 100.0%; Pred. No. 1.4e-41;
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1013 GGGGAGCGGGCGGGATTGGAGACAGAGCCCAAGGCAACAAAGTGGCGGTGAGA 1072
Db 1 GGGGAGCGGGCGGGATTGGAGACAGAGCCCAAGGCAACAAAGTGGCGGTGAGA 60

Qy 1073 AATCAACAGCGGTGCTTCCGAGAGAGAGAGAG 1111
Db 61 AATCAACAGCGGTGCTTCCGAGAGAGAGAGAG 99

RESULT 3
CA593541
LOCUS
DEFINITION wpalc.pk002.p19 wpalc Triticum aestivum cDNA clone wpalc.pk002.p19
5' end, mRNA sequence.
ACCESSION CA593541
VERSION CA593541.1 GI:25143327
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
1 (bases 1 to 594)
REFERENCE
AUTHORS Tingey,S.V., Moore,G., Griffiths,S., Powell,W., Wolters,P.,
Dolan,M., Hainey,C., Miao,G., Caraher,N. and Hanafey,M.K.
TITLE DuPont Wheat cDNA Sequence in collaboration with the John Innes
Center 1
JOURNAL Unpublished (2002)
COMMENT Contact: Scott V. Tingey
Crop Genetics
E. I. DuPont de Nemours and Company
```

```
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
Tel: 302-631-2602
Fax: 302-631-2607
Email: Scott.V.Tingey@USA.dupont.com
Seq primer: M13.
location/Qualifiers
source
1..594
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone="wpalc.pk002.p19"
/tissue_type="anthers"
/lab_host="DH10B"
/clone_lib="wpalc"
/notes="Vector: pBluescript SK+; Site 1: EcoRI; Site 2:
XhoI; Wheat (Triticum aestivum) pre-meiotic anthers_JIC"

ORIGIN
Query Match 8.3%; Score 92; DB 6; Length 594;
Best Local Similarity 100.0%; Pred. No. 8.5e-38;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1020 CGGGCCGGGATTGGAGACAGAGCCCAAGGCAACAAAGTGGCGGTGAGAAATCAAC 1079
Db 1 CGGGCCGGGATTGGAGACAGAGCCCAAGGCAACAAAGTGGCGGTGAGAAATCAAC 60

Qy 1080 AAGCGGTGCTTCCGAGAGAGAGAGAGAG 1111
Db 61 AAGCGGTGCTTCCGAGAGAGAGAGAGAG 92

RESULT 4
CD938039
LOCUS
DEFINITION OV.108007F010206 OV Triticum aestivum cDNA clone OV108007, mRNA
sequence.
ACCESSION CD938039
VERSION CD938039.1 GI:32785547
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
1 (bases 1 to 631)
REFERENCE
AUTHORS Genoplante.
TITLE Genoplante, a major partnership french program in plant genomics
JOURNAL Unpublished (2003)
COMMENT Contact: Genoplante
Genoplante
93, rue Henri Rochefort 91025 EVRY CEDEX France
Tel: 33 1 69 47 54 00
Fax: 33 1 69 47 54 10
This sequence has been generated in the framework of the french
plant genomics programme 'Genoplante' (http://www.genoplante.com
and http://genoplante-info.infobiogen.fr).
location/Qualifiers
source
1..631
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="recital"
/db_xref="taxon:4565"
/clone="OV108007"
/tissue_type="ovary"
/clone_lib="OV"

ORIGIN
Query Match 8.3%; Score 92; DB 6; Length 631;
Best Local Similarity 100.0%; Pred. No. 8.5e-38;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1020 CGGGCCGGGATTGGAGACAGAGCCCAAGGCAACAAAGTGGCGGTGAGAAATCAAC 1079
```

Db 1 CGGCCGGGATTGGACAGAGCCCAAGGCAACAAAGTGGCGTGAGAAATCAAC 60  
Qy 1080 AAGCGGTGCTTGGCGAGAGAGAGAGAG 1111  
Db 61 AAGCGGTGCTTGGCGAGAGAGAGAGAGAG 92

RESULT 5  
CD865240 624 bp mRNA linear EST 11-JUL-2003  
LOCUS AZ02.073123F000912 AZ02 Triticum aestivum cDNA clone AZ02073123,  
DEFINITION mRNA sequence.  
ACCESSION CD865240  
VERSION CD865240.1 GI:32549056  
KEYWORDS EST.  
SOURCE Triticum aestivum (bread wheat)  
ORGANISM Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Pooideae; Triticeae; Triticum.  
REFERENCE 1 (bases 1 to 624)  
AUTHORS Genoplante.  
TITLE Genoplante, a major partnership french program in plant genomics  
JOURNAL Unpublished (2003)  
COMMENT Contact: Genoplante  
Genoplante  
93, rue Henri Rochefort 91025 EVRY CEDEX France  
Tel: 33 1 69 47 54 00  
Fax: 33 1 69 47 54 10  
This sequence has been generated in the framework of the french  
plant genomics programme 'Genoplante' (<http://www.genoplante.com>  
and <http://genoplante-info.infobiogen.fr>).  
FEATURES  
source  
Location/Qualifiers  
1..624  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="recital"  
/db\_xref="taxon:4565"  
/clone="AZ02073123"  
/tissue\_type="root"  
/clone\_lib="AZ02"

ORIGIN  
Query Match 8.1%; Score 90; DB 6; Length 624;  
Best Local Similarity 100.0%; Pred. No. 1e-36;  
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1022 GCGCGGATTGGACAGAGCCCAAGGCAACAAAGTGGCGTGAGAAATCAACAA 1081  
Db 1 GCGCGGATTGGACAGAGCCCAAGGCAACAAAGTGGCGTGAGAAATCAACAA 60

Qy 1082 GCGGTGCTTCCGAGAGAGAGAGAG 1111  
Db 61 GCGGTGCTTCCGAGAGAGAGAGAGAG 90

RESULT 6  
CA712930 430 bp mRNA linear EST 26-NOV-2002  
LOCUS wdk3c.pk008.e19 wdk3c Triticum aestivum cDNA clone wdk3c.pk008.e19  
DEFINITION 5' end, mRNA sequence.  
ACCESSION CA712930  
VERSION CA712930.1 GI:25434723  
KEYWORDS EST.  
SOURCE Triticum aestivum (bread wheat)  
ORGANISM Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Pooideae; Triticeae; Triticum.  
REFERENCE 1 (bases 1 to 430)  
AUTHORS Tingey, S.V., Powell, W., Wolters, P., Dolan, M., Hainey, C., Yuan, Z.,  
Miao, G., Caraher, N. and Hanafey, M.K.  
TITLE DuPont Wheat cDNA Sequence

JOURNAL Unpublished (2002)  
COMMENT Contact: Scott V. Tingey  
Crop Genetics  
E. I. DuPont de Nemours and Company  
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA  
Tel: 302-631-2602  
Fax: 302-631-2607  
Email: Scott.V.Tingey@USA.dupont.com  
Seq primer: M13.

FEATURES  
source  
Location/Qualifiers  
1..430  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/db\_xref="taxon:4565"  
/clone="wdk3c.pk008.e19"  
/tissue\_type="kernel"  
/lab\_host="DH108"  
/clone\_lib="wdk3c"  
/note="Vector: pBluescript SK+; Site 1: EcoRI; Site 2:  
XhoI; Wheat (Triticum aestivum L.) developing kernel, 14  
days after anthesis."  
ORIGIN  
Query Match 7.9%; Score 88; DB 6; Length 430;  
Best Local Similarity 100.0%; Pred. No. 1.2e-35;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1024 CCGGGATTGGACAGAGCCCAAGGCAACAAAGTGGCGTGAGAAATCAACAAAGC 1083  
Db 1 CCGGGATTGGACAGAGCCCAAGGCAACAAAGTGGCGTGAGAAATCAACAAAGC 60

Qy 1084 GTGTCTTCCGAGAGAGAGAGAGAG 1111  
Db 61 GTGTCTTCCGAGAGAGAGAGAGAG 88

RESULT 7  
CD884411 643 bp mRNA linear EST 14-JUL-2003  
LOCUS F1.116120F010507 F1 Triticum aestivum cDNA clone F1116120, mRNA  
DEFINITION sequence.  
ACCESSION CD884411  
VERSION CD884411.1 GI:32648852  
KEYWORDS EST.  
SOURCE Triticum aestivum (bread wheat)  
ORGANISM Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Pooideae; Triticeae; Triticum.  
REFERENCE 1 (bases 1 to 643)  
AUTHORS Genoplante.  
TITLE Genoplante, a major partnership french program in plant genomics  
JOURNAL Unpublished (2003)  
COMMENT Contact: Genoplante  
Genoplante  
93, rue Henri Rochefort 91025 EVRY CEDEX France  
Tel: 33 1 69 47 54 00  
Fax: 33 1 69 47 54 10  
This sequence has been generated in the framework of the french  
plant genomics programme 'Genoplante' (<http://www.genoplante.com>  
and <http://genoplante-info.infobiogen.fr>).  
FEATURES  
source  
Location/Qualifiers  
1..643  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="recital"  
/db\_xref="taxon:4565"  
/clone="F1116120"  
/tissue\_type="leaf one"  
/clone\_lib="F1"

ORIGIN  
Query Match 5.9%; Score 66; DB 6; Length 643;

```

Best Local Similarity 100.0%; Pred. No. 1e-23;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1046 CAAGCGACACAAAGTGGCGGTGAGAAATCAACAAGCGGTGCTTGGCGAGAGAGAGAG 1105
    |||||
Db 1 CAAGCGACACAAAGTGGCGGTGAGAAATCAACAAGCGGTGCTTGGCGAGAGAGAGAG 60
    |||||

Qy 1106 AGAGAG 1111
    |||||
Db 61 AGAGAG 66

RESULT 8
LOCUS CFI132917 642 bp mRNA linear EST 24-JUL-2003
DEFINITION WHE4351_G12_M23T Wheat meiotic floret cDNA library Triticum aestivum CDNA clone WHE4351_G12_M23, mRNA sequence.
ACCESSION CFI132917
VERSION CFI132917.1 GI:33216504
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 642)
AUTHORS Anderson,O.D., Chao,S., Crossman,C., Langridge,P., Lazo,G.R.,
Pham,J., Rausch,C.J., Sutton,T., Woo,J. and Wilson,C.
TITLE The structure and function of the expressed portion of the wheat
genomes - Meiotic floret cDNA library
JOURNAL Unpublished (2003)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105959773
Fax: 5105959818
Email: oandersn@pw.usda.gov
Sequences have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: T7 primer.
FEATURES
Location/Qualifiers
1..642
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE4351_G12_M23"
/tissue_type="Whole florets with anthers"
/dev_stages="Meiotic stages pre-meiosis-metaphase I"
/lab_host="E. coli DH10B"
/clone_lib="Wheat meiotic floret cDNA library"
/note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; Plants
were grown in a glasshouse. Another meiotic stage was
determined microscopically after removing a single anther
from a primary floret. If determined to be between (and
including) meiotic stages pre-meiosis and metaphase I,
remaining floret tissue was collected and pooled for
library construction. The tissue, total RNA, and poly(A)
RNA were prepared, cDNA synthesised, and directionally
ligated into pSPORT1 by Tim Sutton in the P Langridge Lab
at the Department of Plant Science, University of
Adelaide, Waite Campus, Australia. Average insert size
1.4Kb. Plasmid DNA preparations and DNA sequencing were
performed in the OD Anderson lab (all other authors)."

ORIGIN
Query Match 5.8%; Score 64; DB 7; Length 642;
Best Local Similarity 100.0%; Pred. No. 1.2e-22;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1048 AGGCAACAAAGTGGCGGTGAGAAATCAACAAGCGGTGCTTGGCGAGAGAGAGAG 1107
    |||||

```

```

Db 47 AGGCAACAAAGTGGCGGTGAGAAATCAACAAGCGGTGCTTGGCGAGAGAGAGAG 106
    |||||
Qy 1108 AGAG 1111
    |||||
Db 107 AGAG 110

RESULT 9
LOCUS CD894217 712 bp mRNA linear EST 14-JUL-2003
DEFINITION G118.125L08P010828 G118 Triticum aestivum cDNA clone G118125L08,
mRNA sequence.
ACCESSION CD894217
VERSION CD894217.1 GI:32665532
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 712)
AUTHORS Genoplante.
TITLE Genoplante, a major partnership french program in plant genomics
JOURNAL Unpublished (2003)
COMMENT Contact: Genoplante
Genoplante
93, rue Henri Rochefort 91025 EVRY CEDEX France
Tel: 33 1 69 47 54 00
Fax: 33 1 69 47 54 10
This sequence has been generated in the framework of the french
plant genomics programme 'Genoplante' (http://www.genoplante.com
and http://genoplante-info.infobiogen.fr).
FEATURES
Location/Qualifiers
1..712
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="recital"
/db_xref="taxon:4565"
/clone="G118125L08"
/tissue_type="grain (118 degrees per day after
pollination)"
/clone_lib="G118"

ORIGIN
Query Match 5.8%; Score 64; DB 6; Length 712;
Best Local Similarity 100.0%; Pred. No. 1.2e-22;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1048 AGGCAACAAAGTGGCGGTGAGAAATCAACAAGCGGTGCTTGGCGAGAGAGAGAG 1107
    |||||
Db 11 AGGCAACAAAGTGGCGGTGAGAAATCAACAAGCGGTGCTTGGCGAGAGAGAGAG 70
    |||||

Qy 1108 AGAG 1111
    |||||
Db 71 AGAG 74

RESULT 10
LOCUS CA701748 556 bp mRNA linear EST 26-NOV-2002
DEFINITION wkm2c.pk005.j20 wkm2c Triticum aestivum cDNA clone wkm2c.pk005.j20
5' end, mRNA sequence.
ACCESSION CA701748
VERSION CA701748.1 GI:25423541
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 556)
AUTHORS Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z.,
Miao,G., Caraher,N. and Hanafey,M.K.

```



```

TITLE      DuPont Wheat cDNA Sequence
JOURNAL    Unpublished (2002)
COMMENT    Contact: Scott V. Tingey
           Crop Genetics
           E. I. DuPont de Nemours and Company
           1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
           Tel: 302-631-2602
           Fax: 302-631-2607
           Email: Scott.V.Tingey@USA.dupont.com
           Seq primer: M13.

FEATURES   Location/Qualifiers
source     1..556
           /organism="Triticum aestivum"
           /mol_type="mRNA"
           /cultivar="hard red spring"
           /db_xref="taxon:4565"
           /clone="wkm2c.pk005.j20"
           /tissue_type="kernel"
           /lab_host="DH10B"
           /clone_lib="wkm2c"
           /note="Site 1: EcoRI; Site 2: XhoI; Wheat (Triticum
           aestivum L.) kernel malted 175 hours at 4 C"

ORIGIN
Query Match      5.7%; Score 63; DB 6; Length 556;
Best Local Similarity 100.0%; Pred. No. 4.2e-22;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1049 GGCACACAAACGTCGCGTGAGAAATCAACACAGCGGTCTGCCGAGAGAGAGAGA 1108
      |||||
Db 1 GGCACACAAACGTCGCGTGAGAAATCAACACAGCGGTCTGCCGAGAGAGAGAGA 60

QY 1109 GAG 1111
      |||
Db 61 GAG 63

RESULT 11
CD904051      637 bp mRNA linear EST 14-JUL-2003
LOCUS      G356.112E12F010920 G356 Triticum aestivum cDNA clone G356112E12,
DEFINITION mRNA sequence.
ACCESSION      CD904051
VERSION      CD904051.1 GI:32678379
KEYWORDS      EST.
SOURCE      Triticum aestivum (bread wheat)
ORGANISM      Triticum aestivum

REFERENCE
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Triticum.
1 (bases 1 to 637)
Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T.,
Cloutier, S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A.,
Gustafson, P., Herrmann, R.G., Holton, T., Jacquemin, J.M., Jia, J.,
Joudrier, P., Langridge, P., Laz, G.R., Lin, J.J., McGuire, P.,
Ogihara, Y., Pecchioni, N., Qualset, C., Schuch, W., Selvaraj, G.,
Shariflou, M., Sorrells, M., Warburton, M., and Wenzel, G.
International Triticeae EST Cooperative (ITEC): Production of
Expressed Sequence Tags for Species of the Triticeae
Unpublished (2000)
Contact: Schuch W
Zeneca Wheat Improvement Centre, Norwich Research Park
Colney Lane, Norwich NR4 7UH UNITED KINGDOM
Tel: 44 1603 250 2600
Fax: 44 1603 250 699
Email: wolfgang.schuch@aguk.zeneca.com
International Triticeae EST Cooperative (ITEC)
http://wheat.pw.usda.gov/genome.

FEATURES   Location/Qualifiers
source     1..422
           /organism="Triticum aestivum"
           /mol_type="mRNA"
           /cultivar="Novosibirskaya 67"
           /db_xref="taxon:4565"
           /clone="WWR021.H3"
           /tissue_type="root"
           /clone_lib="ITEC WMR Wheat Root Library"
           /note="M13 Reverse sequencing primer used for 5' end of
           clone."

ORIGIN
Query Match      5.6%; Score 62; DB 2; Length 422;
Best Local Similarity 100.0%; Pred. No. 1.5e-21;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1050 GCAACACAAAGTCGCGTGAGAAATCAACACAGCGGTCTGCCGAGAGAGAGAGAG 1109
      |||||
Db 8 GCAACACAAAGTCGCGTGAGAAATCAACACAGCGGTCTGCCGAGAGAGAGAGAG 67

QY 1110 AG 1111
      ||
Db 68 AG 69

RESULT 13
BJ257084

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 30, 2005, 09:49:55 ; Search time 5145 Seconds  
(without alignments)  
10463.313 Million cell updates/sec

Title: US-09-979-549-2

Perfect score: 1111

Sequence: 1 gaagtcagaagccgttcag.....cogagaagagagagagag 1111

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1111	100.0	1111	6	BD263890
2	1111	100.0	1111	6	AX047682 Sequence
3	1111	100.0	2687	6	BD263889 Promoter
4	1111	100.0	2687	6	AX047681 Sequence
5	71.2	6.4	157969	8	AP005428 Oryza sat
6	68.6	6.2	629	8	TAE0404845
7	65.2	5.9	154198	8	AP005524 Oryza sat
8	65.2	5.9	159049	8	AP005544 Oryza sat
9	62.8	5.7	140715	8	AB109206 Oryza sat
10	62.8	5.7	169030	2	AP005970 Oryza sat
11	62.8	5.7	171257	8	AP006149 Oryza sat
12	62.6	5.6	152423	8	AC079853 Oryza sat
13	62.6	5.6	169728	8	AC091787 Oryza sat
14	59.4	5.3	127901	8	AC121366 Oryza sat
15	59.4	5.3	135231	8	AC137618 Oryza sat
16	59.4	5.3	146435	8	AC135927 Oryza sat
17	58.4	5.3	161563	8	AP003076 Oryza sat
18	58.4	5.3	162776	8	AP003237 Oryza sat
19	57.8	5.2	2000	6	AX655816 Sequence

C 20	56.8	5.1	132628	8	AP005517	Oryza sat
C 21	56.8	5.1	148527	8	AP005410	Oryza sat
C 22	56.6	5.1	148892	8	AP003235	Oryza sat
C 23	56.6	5.1	152274	8	AP004669	Oryza sat
C 24	56.6	5.1	152736	8	AP003566	Oryza sat
C 25	56	5.0	142752	8	AC093953	Oryza sat
C 26	56	5.0	160938	8	AC119670	Oryza sat
C 27	55.6	5.0	148707	8	AP002745	Oryza sat
C 28	55.6	5.0	148985	8	AP002094	Oryza sat
C 29	54.4	4.9	152240	8	AP003254	Oryza sat
C 30	54	4.9	596	8	TAE9762	Triticum
C 31	52.8	4.8	155106	8	AC097174	Oryza sat
C 32	52.8	4.8	160142	8	AC147462	Oryza sat
C 33	52.8	4.8	185623	8	AC137611	Oryza sat
C 34	51	4.6	139788	8	AC145380	Oryza sat
C 35	50.6	4.6	147638	8	AP003303	Oryza sat
C 36	50.6	4.6	152556	8	AP003216	Oryza sat
C 37	50.4	4.5	203132	8	AP004071	Oryza sat
C 38	49.4	4.4	107407	8	AC134933	Oryza sat
C 39	49	4.4	149155	8	AP005113	Oryza sat
C 40	49	4.4	158719	8	AP005619	Oryza sat
C 41	48	4.3	106166	8	AC105769	Oryza sat
C 42	47.6	4.3	2591	8	AK106639	Oryza sat
C 43	47.6	4.3	159861	8	AP005012	Oryza sat
C 44	47.4	4.3	101104	8	AP003370	Oryza sat
C 45	47.4	4.3	150066	8	AP003346	Oryza sat

ALIGNMENTS

RESULT 1	BD263890	1111 bp	DNA	linear	PAT 17-JUL-2003
LOCUS	Promoter of thioredoxine TaTrxh2 in wheat.				
DEFINITION	BD263890				
ACCESSION	JP 2002543844-A/2.				
VERSION	JP 2002543844-A/2.				
KEYWORDS	Triticum aestivum (bread wheat)				
SOURCE	Triticum aestivum				
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.				
REFERENCE	1 (bases 1 to 1111)				
AUTHORS	Gautier,M.F., Ithorai,T. and Joudrier,P.				
TITLE	Promoter of thioredoxine TaTrxh2 in wheat				
JOURNAL	Patent: JP 2002543844-A 2 24-DEC-2002;				
COMMENT	INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE				
	OS Triticum aestivum (common wheat)				
	PN JP 2002543844-A/2				
	PD 24-DEC-2002				
	PF 17-MAY-2000 JP 2000618471				
	PR 17-MAY-1999 FR 99/06231				
	PI MARIE FRANCOISE GAUTIER,TANIA ITHORAI,PHILIPPE JOUDRIER PC				
	C12N15/09,A01H5/00,C12N5/10// (C12N5/10,C12R1:91),C12N15/00, PC				
	C12N5/00,				
	PC (C12N5/00,C12R1:91)				
	CC Promoter of thioredoxine TaTrxh2 in wheat				
	PH Key Location/Qualifiers				
	FT source				
	FT i. .1111				
	Location/Qualifiers				
	1. .1111				
	/organism="Triticum aestivum"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:4565"				

FEATURES

source

ORIGIN

Query Match 100.0%; Score 1111; DB 6; Length 1111;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GAAGTCAGAAGCCGTTTCAAGATTGTTGGAGGACTCGAAAAAAGAGGGAGCCCGGC 60

```
Db 1 GAAAGTCAGAGGCGGTTTTCAGAAATGTTGGAGGACTCGAAAAAAGAGGGGAGCCAGGC 60
Qy 61 AGACGAGGGGCGCATGTGCTTCTTGGCGAGGCGTCTAGCTTTGGCAGCGCGGC 120
Db 61 AGACGAGGGGCGCATGTGCTTCTTGGCGAGGCGTCTAGCTTTGGCAGCGCGGC 120
Qy 121 CGCTTTTCTCTTGGGTGGGCGCGGAGCTCCCGAGTTTGGCGCAATTTTTCAT 180
Db 121 CGCTTTTCTCTTGGGTGGGCGCGGAGCTCCCGAGTTTGGCGCAATTTTTCAT 180
Qy 181 TTTATGCGCATGGCGTCAAGCGTTTATCTAGGCGTCTGGAGGGTACATTTGAAGATGTG 240
Db 181 TTTATGCGCATGGCGTCAAGCGTTTATCTAGGCGTCTGGAGGGTACATTTGAAGATGTG 240
Qy 241 CCACCAACTCCAAACCGACACCTGTATCTGAGCATGCCCTCATGCTCTCTTCATGCC 300
Db 241 CCACCAACTCCAAACCGACACCTGTATCTGAGCATGCCCTCATGCTCTCTTCATGCC 300
Qy 301 TCCCTTTGGGTGAGGTCAATGTGCCCTTGGGCGGAGTGGCTTCCCGTTTGAAGCAAGTAT 360
Db 301 TCCCTTTGGGTGAGGTCAATGTGCCCTTGGGCGGAGTGGCTTCCCGTTTGAAGCAAGTAT 360
Qy 361 AATAAGTCTTAGTCAGCTGGCTATAAGATGTTCCACATCAGCAAAATCCTTAAACTGAGG 420
Db 361 AATAAGTCTTAGTCAGCTGGCTATAAGATGTTCCACATCAGCAAAATCCTTAAACTGAGG 420
Qy 421 AGAAGAAAGTAGAGTGAGAGGGCGTCCGCGTCTCGTCAATCGCTAGGATAGCACAA 480
Db 421 AGAAGAAAGTAGAGTGAGAGGGCGTCCGCGTCTCGTCAATCGCTAGGATAGCACAA 480
Qy 481 GCTCCCACTGGAATCGAGCAACATGCAACCGCACATGACTAAAGGCAAAACGCCCA 540
Db 481 GCTCCCACTGGAATCGAGCAACATGCAACCGCACATGACTAAAGGCAAAACGCCCA 540
Qy 541 ATCAGTATGCTTTCTCTGCAATCTTTCTCATGCAAGCAATTAATCTATAGCTAAATCTA 600
Db 541 ATCAGTATGCTTTCTCTGCAATCTTTCTCATGCAAGCAATTAATCTATAGCTAAATCTA 600
Qy 601 CAGCAGTTTATTAATAACAGGCTATATAGCTGACTGGCAGTGTATAGAGCGCGCA 660
Db 601 CAGCAGTTTATTAATAACAGGCTATATAGCTGACTGGCAGTGTATAGAGCGCGCA 660
Qy 661 GCCGCTCTTCTATTAGCTTTGCTCTTATGCTATCATCTGTGTGAGCAGTCAATGATTC 720
Db 661 GCCGCTCTTCTATTAGCTTTGCTCTTATGCTATCATCTGTGTGAGCAGTCAATGATTC 720
Qy 721 AAACAACAAATCCGGGCGTTGAGCAAGTCGGAATGAATTTCCGCTCATCTATTTGTCG 780
Db 721 AAACAACAAATCCGGGCGTTGAGCAAGTCGGAATGAATTTCCGCTCATCTATTTGTCG 780
Qy 781 TGGGCTCACGCTATTCGCTTAACCGGTTTGAATCAGACCTTCAGAGCCACGCGCTC 840
Db 781 TGGGCTCACGCTATTCGCTTAACCGGTTTGAATCAGACCTTCAGAGCCACGCGCTC 840
Qy 841 CAGGACCGGTTTACCGCTCAGCCTAAAAAAGAAAAAATCTGTTCATCATCACGCC 900
Db 841 CAGGACCGGTTTACCGCTCAGCCTAAAAAAGAAAAAATCTGTTCATCATCACGCC 900
Qy 901 CATCTGAACCGTTCAAGAGCCCAAGTAAATTTCCGGCACGAGCAAGGGCATATCCGTCA 960
Db 901 CATCTGAACCGTTCAAGAGCCCAAGTAAATTTCCGGCACGAGCAAGGGCATATCCGTCA 960
Qy 961 TAGCGAGCGCATAAATCTGATCTGCTGCCGCAATTTATCTTTGGGAGGC 1020
Db 961 TAGCGAGCGCATAAATCTGATCTGCTGCCGCAATTTATCTTTGGGAGGC 1020
Qy 1021 GGGCGGGAATGGAGACAGAGCCCAAGGCAACAAAGTGGCGGTGAGAAATCAACA 1080
Db 1021 GGGCGGGAATGGAGACAGAGCCCAAGGCAACAAAGTGGCGGTGAGAAATCAACA 1080
Qy 1081 AGCGGTGCTTCCGAGAGAGAGAGAG 1111
```

```
Db 1081 AGCGGTGCTTCCGAGAGAGAGAGAGAG 1111
RESULT 2
AX047682 1111 bp DNA linear PAT 15-DEC-2000
LOCUS Sequence 2 from Patent WO0070065.
DEFINITION AX047682
ACCESSION AX047682
VERSION AX047682.1 GI:11876717
KEYWORDS Triticum aestivum (bread wheat)
SOURCE Triticum aestivum
ORGANISM Triticum aestivum
REFERENCE 1. 1111
AUTHORS Gautier,M.F., Ithorai,T. and Joudrier,P.
TITLE Promoter of thioedoxine tatrzh2 in wheat
JOURNAL Patent: WO 0070065-A 2 23-NOV-2000;
INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (INRA) (FR)
FEATURES
source
1. 1111
/organism="Triticum aestivum"
/mol_type="unassigned DNA"
/db_xref="taxon:4565"
ORIGIN
Query Match 100.0%; Score 1111; DB 6; Length 1111;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GAAATCAGAGGCGGTTTTCAGAAATGTTGGAGGACTCGAAAAAAGAGGGGAGCCAGGC 60
Db 1 GAAATCAGAGGCGGTTTTCAGAAATGTTGGAGGACTCGAAAAAAGAGGGGAGCCAGGC 60
Qy 61 AGACGAGGGGCGCATGTGCTTCTTGGCGAGGCGTCTAGCTTTGGCAGCGCGGC 120
Db 61 AGACGAGGGGCGCATGTGCTTCTTGGCGAGGCGTCTAGCTTTGGCAGCGCGGC 120
Qy 121 CGCTTTTCTCTTGGGTGGGCGCGGAGTGGCTTCCCTTGGCGAGGCGTCTAGCTTTGGCAGCGCGGC 180
Db 121 CGCTTTTCTCTTGGGTGGGCGCGGAGTGGCTTCCCTTGGCGAGGCGTCTAGCTTTGGCAGCGCGGC 180
Qy 181 TTTATGCGCATGGCGTCAAGCGTTTATCTAGGCGTCTGGAGGGTACATTTGAAGATGTG 240
Db 181 TTTATGCGCATGGCGTCAAGCGTTTATCTAGGCGTCTGGAGGGTACATTTGAAGATGTG 240
Qy 241 CCACCAACTCCAAACCGACACCTGTATCTGAGCATGCCCTCATGCTCTCTTCATGCC 300
Db 241 CCACCAACTCCAAACCGACACCTGTATCTGAGCATGCCCTCATGCTCTCTTCATGCC 300
Qy 301 TCCCTTTGGGTGAGGTCAATGTGCCCTTGGGCGGAGTGGCTTCCCGTTTGAAGCAAGTAT 360
Db 301 TCCCTTTGGGTGAGGTCAATGTGCCCTTGGGCGGAGTGGCTTCCCGTTTGAAGCAAGTAT 360
Qy 361 AATAAGTCTTAGTCAGCTGGCTATAAGATGTTCCACATCAGCAAAATCCTTAAACTGAGG 420
Db 361 AATAAGTCTTAGTCAGCTGGCTATAAGATGTTCCACATCAGCAAAATCCTTAAACTGAGG 420
Qy 421 AGAAGAAAGTAGAGTGAGAGGGCGTCCGCGTCTCGTCAATCGCTAGGATAGCACAA 480
Db 421 AGAAGAAAGTAGAGTGAGAGGGCGTCCGCGTCTCGTCAATCGCTAGGATAGCACAA 480
Qy 481 GCTCCCACTGGAATCGAGCAACATGCAACCGCACATGACTAAAGGCAAAACGCCCA 540
Db 481 GCTCCCACTGGAATCGAGCAACATGCAACCGCACATGACTAAAGGCAAAACGCCCA 540
Qy 541 ATCAGTATGCTTTCTCTGCAATCTTTCTCATGCAAGCAATTAATCTATAGCTAAATCTA 600
Db 541 ATCAGTATGCTTTCTCTGCAATCTTTCTCATGCAAGCAATTAATCTATAGCTAAATCTA 600
Qy 601 CAGCAGTTTATTAATAACAGGCTATATAGCTGACTGGCAGTGTATAGAGCGCGCA 660
Db 601 CAGCAGTTTATTAATAACAGGCTATATAGCTGACTGGCAGTGTATAGAGCGCGCA 660
Qy 661 GCCGCTCTTCTATTAGCTTTGCTCTTATGCTATCATCTGTGTGAGCAGTCAATGATTC 720
Db 661 GCCGCTCTTCTATTAGCTTTGCTCTTATGCTATCATCTGTGTGAGCAGTCAATGATTC 720
Qy 721 AAACAACAAATCCGGGCGTTGAGCAAGTCGGAATGAATTTCCGCTCATCTATTTGTCG 780
Db 721 AAACAACAAATCCGGGCGTTGAGCAAGTCGGAATGAATTTCCGCTCATCTATTTGTCG 780
Qy 781 TGGGCTCACGCTATTCGCTTAACCGGTTTGAATCAGACCTTCAGAGCCACGCGCTC 840
Db 781 TGGGCTCACGCTATTCGCTTAACCGGTTTGAATCAGACCTTCAGAGCCACGCGCTC 840
Qy 841 CAGGACCGGTTTACCGCTCAGCCTAAAAAAGAAAAAATCTGTTCATCATCACGCC 900
Db 841 CAGGACCGGTTTACCGCTCAGCCTAAAAAAGAAAAAATCTGTTCATCATCACGCC 900
Qy 901 CATCTGAACCGTTCAAGAGCCCAAGTAAATTTCCGGCACGAGCAAGGGCATATCCGTCA 960
Db 901 CATCTGAACCGTTCAAGAGCCCAAGTAAATTTCCGGCACGAGCAAGGGCATATCCGTCA 960
Qy 961 TAGCGAGCGCATAAATCTGATCTGCTGCCGCAATTTATCTTTGGGAGGC 1020
Db 961 TAGCGAGCGCATAAATCTGATCTGCTGCCGCAATTTATCTTTGGGAGGC 1020
Qy 1021 GGGCGGGAATGGAGACAGAGCCCAAGGCAACAAAGTGGCGGTGAGAAATCAACA 1080
Db 1021 GGGCGGGAATGGAGACAGAGCCCAAGGCAACAAAGTGGCGGTGAGAAATCAACA 1080
Qy 1081 AGCGGTGCTTCCGAGAGAGAGAGAGAG 1111
```

Db	601	CAGCGAGTTTATTATATAAAGAGGCTATATAGCTGACCTGGCAGTGTCTATAGAGCCGGCA	660
Qy	661	GCGGGCTCTTCTATTAGCTTTGCTCTTATGGCTACATCTGTGAGCAGTCGATGATTC	720
Db	661	GCGGGCTCTTCTATTAGCTTTGCTCTTATGGCTACATCTGTGAGCAGTCGATGATTC	720
Qy	721	AAACAACAATCCGGGCGTTTACGAGTCGGAATGAATTTGGGCTCATCTCATTTGTGCG	780
Db	721	AAACAACAATCCGGGCGTTTACGAGTCGGAATGAATTTGGGCTCATCTCATTTGTGCG	780
Qy	781	TGGGCTCACGCGTATTTCGCTTACCGTGTGGAATCAGACCTCTCAGGAAGCCACGGCTC	840
Db	781	TGGGCTCACGCGTATTTCGCTTACCGTGTGGAATCAGACCTCTCAGGAAGCCACGGCTC	840
Qy	841	CAGCGACCGGTTTACCAACGTCAGCTTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	900
Db	841	CAGCGACCGGTTTACCAACGTCAGCTTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	900
Qy	901	CATCTGAACCGTTTCAACAGCCGCAAGTAAATTTGCGGACCCAGCAAGGAGGATTCGCTCA	960
Db	901	CATCTGAACCGTTTCAACAGCCGCAAGTAAATTTGCGGACCCAGCAAGGAGGATTCGCTCA	960
Qy	961	TAGCGAGCGATTAATTTCTGATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1020
Db	961	TAGCGAGCGATTAATTTCTGATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1020
Qy	1021	GGGCGGGGTTGGAGACAGAGCCACAAAGGCAACAAAGGTCGCTGAGAAATCAACA	1080
Db	1021	GGGCGGGGTTGGAGACAGAGCCACAAAGGCAACAAAGGTCGCTGAGAAATCAACA	1080
Qy	1081	AGCGGCTGCTTCCGAGAGAGAGAGAGAG 1111	
Db	1081	AGCGGCTGCTTCCGAGAGAGAGAGAGAG 1111	
RESULT 3			
BD263889			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum.			
1 (bases 1 to 2687)			
Gautier, M.F., Ihorai, T. and Joudrier, P.			
Promoter of thioredoxine TaTrxh2 in wheat			
Patent: JP 2002543844-A 1 24-DEC-2002;			
INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE			
OS Triticum aestivum (common wheat)			
PN JP 2002543844-A/1			
PD 24-DEC-2002			
PF 17-MAY-2000 JP 2000618471			
PR 17-MAY-1999 FR 95/06231			
PI MARIE FRANCOISE GAUTIER, TANITA IHORAI, PHILIPPE JOUDRIER PC			
C12N15/09, A01H5/00, C12N5/10, C12N5/10, C12R1/91, C12N15/00, PC			
C12N5/00,			
PC (C12N15/00, C12R1/91)			
CC Promoter of thioredoxine TaTrxh2 in wheat			
FH Key Location/Qualifiers			
FT exon (1112).. (1231)			
FT intron (1232).. (2203)			
FT exon (2204).. (2326)			
FT intron (2327).. (2420)			
FT exon (2421).. (2558)			
FT CDS (1112).. (1231)			
FT CDS (2204).. (2326)			
FT CDS (2421).. (2558)			
Location/Qualifiers			
1..2687			
FEATURES			
source			

/organism="Triticum aestivum"			
/mol_type="genomic DNA"			
/db_xref="taxon:4565"			
ORIGIN			
Query Match 100.0%; Score 1111; DB 6; Length 2687;			
Best Local Similarity 100.0%; Pred. No. 0;			
Matches 1111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	GAAGTCAGAAAGCCGCTTCAGAAATGTTGGAGGACTCGAAAGGAGGAGGAGGAGGAGGAGGAGG	60
Db	1	GAAGTCAGAAAGCCGCTTCAGAAATGTTGGAGGACTCGAAAGGAGGAGGAGGAGGAGGAGGAGG	60
Qy	61	AGACGAGCGGGCGGCATGTCCTGTTCTTGGGAGGCGCTCTAGCTTTGGGAGGCGGCGG	120
Db	61	AGACGAGCGGGCGGCATGTCCTGTTCTTGGGAGGCGCTCTAGCTTTGGGAGGCGGCGG	120
Qy	121	CGCTTTTCTCTTGGTGGGCGGCGGAGCTCCCGAGTTTGAGCCGCAATTTTTTACAT	180
Db	121	CGCTTTTCTCTTGGTGGGCGGCGGAGCTCCCGAGTTTGAGCCGCAATTTTTTACAT	180
Qy	181	TTTATGGCGATGGCGTCAGGCGTTTATCTAGGCGTCTGGGAGGTTACATTTGAAGATGTG	240
Db	181	TTTATGGCGATGGCGTCAGGCGTTTATCTAGGCGTCTGGGAGGTTACATTTGAAGATGTG	240
Qy	241	CCACCAATCTCCAAACCGCAACCCCTGTATCTGAGCATGCTCATGCTCTCTCTTCATGCC	300
Db	241	CCACCAATCTCCAAACCGCAACCCCTGTATCTGAGCATGCTCATGCTCTCTCTTCATGCC	300
Qy	301	TCCCTTTGGTGAGTTCATGTCCTTGGGCGGAGTGGCTTCCCGTTTAGAGCAAGTAT	360
Db	301	TCCCTTTGGTGAGTTCATGTCCTTGGGCGGAGTGGCTTCCCGTTTAGAGCAAGTAT	360
Qy	361	AATAAGTCTTAGTCAGCTGGCTATAAGATGTTCCACATCAGCAAAATCTTTAAATCGAGG	420
Db	361	AATAAGTCTTAGTCAGCTGGCTATAAGATGTTCCACATCAGCAAAATCTTTAAATCGAGG	420
Qy	421	AGAAAGAAAGTAGGAGTGAAGAGGCGTTCGGCGCTTCGTCAATCGCTAGCGATAGCACAA	480
Db	421	AGAAAGAAAGTAGGAGTGAAGAGGCGTTCGGCGCTTCGTCAATCGCTAGCGATAGCACAA	480
Qy	481	GCTCCCATGGAATCGAGCCCAACATGCAACCCGACACATGACTAAAGGCAAAACGCCAGCCA	540
Db	481	GCTCCCATGGAATCGAGCCCAACATGCAACCCGACACATGACTAAAGGCAAAACGCCAGCCA	540
Qy	541	ATCAGTATGCTTCTCTGTCATCTTCTTCATGCAAGCATTAATACTATAGCTAATCTA	600
Db	541	ATCAGTATGCTTCTCTGTCATCTTCTTCATGCAAGCATTAATACTATAGCTAATCTA	600
Qy	601	CAGCCAGTTTATTATATAAAGAGGCTTATAGCTGACCTGGCAGTGTCTATAGAGCCGGCA	660
Db	601	CAGCCAGTTTATTATATAAAGAGGCTTATAGCTGACCTGGCAGTGTCTATAGAGCCGGCA	660
Qy	661	GCGGCTCTTCTATTAGCTTTGCTCTTATGGCTACATCTGTGTGAGCAGTTCGATTC	720
Db	661	GCGGCTCTTCTATTAGCTTTGCTCTTATGGCTACATCTGTGTGAGCAGTTCGATTC	720
Qy	721	AAACAACAATCCGGGCGTTTACGAAAGTGAATTTGGGCTCATCTCACTATTGTGCG	780
Db	721	AAACAACAATCCGGGCGTTTACGAAAGTGAATTTGGGCTCATCTCACTATTGTGCG	780
Qy	781	TGGGCTCACGCGTATTTCGCTTAAACCGTTTTCGCTTAAACCGCTTTCGAAATCAGACCTC	840
Db	781	TGGGCTCACGCGTATTTCGCTTAAACCGTTTTCGCTTAAACCGCTTTCGAAATCAGACCTC	840
Qy	841	CAGCGACCGGTTTACCAACGTCAGCTTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	900
Db	841	CAGCGACCGGTTTACCAACGTCAGCTTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	900
Qy	901	CATCTGAACCGTTTCAACAGCCGCAAGTAAATTTGCGGACCCAGCAAGGAGGATTCGCTCA	960
Db	901	CATCTGAACCGTTTCAACAGCCGCAAGTAAATTTGCGGACCCAGCAAGGAGGATTCGCTCA	960

Qy 961 TAGCGAGCGCATAAATCTGATTCCTGCTGCTGCGGCAATATTTATCTTTGGGAGGC 1020  
Db 961 TAGCGAGCGCATAAATCTGATTCCTGCTGCTGCGGCAATATTTATCTTTGGGAGGC 1020  
Qy 1021 GGGCCGGGATTTGGAGACAGAGCCACAAAGGCAACAAAGATGCGCTGAGAAATCAACA 1080  
Db 1021 GGGCCGGGATTTGGAGACAGAGCCACAAAGGCAACAAAGATGCGCTGAGAAATCAACA 1080  
Qy 1081 AGCGGTCTTCCGAGAGAGAGAGAGAG 1111  
Db 1081 AGCGGTCTTCCGAGAGAGAGAGAGAGAG 1111  
RESULT 4  
LOCUS AX047681 2687 bp DNA linear PAT 15-DEC-2000  
DEFINITION Sequence 1 from Patent WO0070065.  
ACCESSION AX047681  
VERSION AX047681.1 GI:11876716  
KEYWORDS  
SOURCE Triticum aestivum (bread wheat)  
ORGANISM Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Triticum.  
REFERENCE 1  
AUTHORS Gautier, M.F., Iborai, T. and Joudrier, P.  
TITLE Promoter of thioresoxine tatr2h2 in wheat  
JOURNAL Patent: WO 0070065-A 1 23-NOV-2000;  
INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (INRA) (FR)  
FEATURES  
source 1..2687  
/organism="Triticum aestivum"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:4565"  
1112..1231  
exon 1232..2203  
intron 2204..2326  
exon 2327..2420  
intron 2421..2558  
exon  
ORIGIN  
Query Match 100.0%; Score 1111; DB 6; Length 2687;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GAAGTCAGAGCGCGTTTCAGAAATGTTGGAGACTCGAAAAAGAGAGGCGCCAGGC 60  
Db 1 GAAGTCAGAGCGCGTTTCAGAAATGTTGGAGACTCGAAAAAGAGAGGCGCCAGGC 60  
Qy 61 AGACGACGGGCGGCATGCTGCTGTTCTTGGCAGGCGCTAGCTTTGGCAGCGCGCG 120  
Db 61 AGACGACGGGCGGCATGCTGCTGTTCTTGGCAGGCGCTAGCTTTGGCAGCGCGCG 120  
Qy 121 CGCTTTTCTCTTGGTGGGCGCGAGCTCCCGAGTTTTCAGCGCAATTTTTCAT 180  
Db 121 CGCTTTTCTCTTGGTGGGCGCGAGCTCCCGAGTTTTCAGCGCAATTTTTCAT 180  
Qy 181 TTTATGGCGATGGCGTTCAGGCGTTTATCTAGGCGTCTGGAGGCTACATTTGAAGATGTG 240  
Db 181 TTTATGGCGATGGCGTTCAGGCGTTTATCTAGGCGTCTGGAGGCTACATTTGAAGATGTG 240  
Qy 241 CCACCAACTCCAAACCGCAACCCCTGTATCTGAGCATGCCCTCATGCCCTCTCTTCATGCC 300  
Db 241 CCACCAACTCCAAACCGCAACCCCTGTATCTGAGCATGCCCTCATGCCCTCTCTTCATGCC 300  
Qy 301 TCCCTTTGGTGAGTTCATGTCCTTTCGGCGCGAGTGGCTTCCCGTTTTCAGCAAGTAT 360  
Db 301 TCCCTTTGGTGAGTTCATGTCCTTTCGGCGCGAGTGGCTTCCCGTTTTCAGCAAGTAT 360  
Qy 361 AATAGTCTTCTAGTCTAGTCTTATAGATGTTCCACATCAGCAATCTTAACTGGAGG 420  
Db 361 AATAGTCTTCTAGTCTAGTCTTATAGATGTTCCACATCAGCAATCTTAACTGGAGG 420

Qy 421 AGAAAGAAAGTAGAGTGAAGAGGGCGTTCGGTCAATCGCTAGCGATAGCACAA 480  
Db 421 AGAAAGAAAGTAGAGTGAAGAGGGCGTTCGGTCAATCGCTAGCGATAGCACAA 480  
Qy 481 GTTCCCATGGAATCGAGCCAAACATGCAACCCGCAATGACTAAAGGCAACCCGAGCCA 540  
Db 481 GTTCCCATGGAATCGAGCCAAACATGCAACCCGCAATGACTAAAGGCAACCCGAGCCA 540  
Qy 541 ATCAGTATGCTTCTCTGTCATCTTCTTCAATGCAAGCAATTAATACTATAGTATCTA 600  
Db 541 ATCAGTATGCTTCTCTGTCATCTTCTTCAATGCAAGCAATTAATACTATAGTATCTA 600  
Qy 601 CAGCCAGTTTATATATAAAGCAGCTATATAGCTGACCTGGCAGTCTATAGAGCCGGCA 660  
Db 601 CAGCCAGTTTATATATAAAGCAGCTATATAGCTGACCTGGCAGTCTATAGAGCCGGCA 660  
Qy 661 GCGGCTCTTCTATAGCTTTCCTTATAGGCTACATCTGTGTGAGCAGTTCGATTGATTC 720  
Db 661 GCGGCTCTTCTATAGCTTTCCTTATAGGCTACATCTGTGTGAGCAGTTCGATTGATTC 720  
Qy 721 AAACAACAATCCGGCGTTTCAGCAAGTCGAAATGAATTTCCGGCTCATCTCATTTGTCG 780  
Db 721 AAACAACAATCCGGCGTTTCAGCAAGTCGAAATGAATTTCCGGCTCATCTCATTTGTCG 780  
Qy 781 TGGGCTCTCAGCGTATTTCCGCTAACCGTGTTTGAATCAGACCTCAGAACCCAGCGCTC 840  
Db 781 TGGGCTCTCAGCGTATTTCCGCTAACCGTGTTTGAATCAGACCTCAGAACCCAGCGCTC 840  
Qy 841 CAGGACCCGTTTTCAGCAGTCCAGCTTAAAAAGAAAAAATACTTTCAATCACAGCC 900  
Db 841 CAGGACCCGTTTTCAGCAGTCCAGCTTAAAAAGAAAAAATACTTTCAATCACAGCC 900  
Qy 901 CATCTGAACCGTTTCAACAGCCCGCAGCTAATTTCCGCGCAGCAAGAGGGGATATCCGTC 960  
Db 901 CATCTGAACCGTTTCAACAGCCCGCAGCTAATTTCCGCGCAGCAAGAGGGGATATCCGTC 960  
Qy 961 TAGCGAGCGCATAAATCTGATTCCTGCTGCTGCGGCAATTTATCTTTGGGAGGC 1020  
Db 961 TAGCGAGCGCATAAATCTGATTCCTGCTGCTGCGGCAATTTATCTTTGGGAGGC 1020  
Qy 1021 GGGCCGGGATTTGGAGACAGAGCCACAAAGGCAACAAAGATGCGCTGAGAAATCAACA 1080  
Db 1021 GGGCCGGGATTTGGAGACAGAGCCACAAAGGCAACAAAGATGCGCTGAGAAATCAACA 1080  
Qy 1081 AGCGGTCTTCCGAGAGAGAGAGAGAGAG 1111  
Db 1081 AGCGGTCTTCCGAGAGAGAGAGAGAGAG 1111  
RESULT 5  
LOCUS AP005428 157969 bp DNA linear PLN 13-JUL-2004  
DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 2,  
PAC clone: P0693E08.  
ACCESSION AP005428  
VERSION AP005428.2 GI:45382012  
KEYWORDS  
SOURCE Oryza sativa (japonica cultivar-group)  
ORGANISM Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartioideae; Oryzaceae; Oryza.  
REFERENCE 1  
AUTHORS Sasaki, T., Matsumoto, T. and Katayose, Y.  
TITLE Oryza sativa nipponbare (GA3) genomic DNA, chromosome 2, PAC  
clone: P0693E08  
JOURNAL Published Only in Database (2002)  
REFERENCE 2 (bases 1 to 157969)  
AUTHORS Sasaki, T., Matsumoto, T. and Katayose, Y.  
TITLE Direct Submission  
JOURNAL Submitted (12-JUN-2002) Takuji Sasaki, National Institute of  
Agrobiological Sciences, Rice Genome Research Program; Kannondai

2-1-2, Teukuba, Ibaraki 305-8602, Japan  
(E-mail:tsaaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,  
Tel:81-298-38-7441, Fax:81-298-38-7468)  
On Mar 11, 2004 this sequence version replaced gi:21396538.  
Genes were predicted from the integrated results of the following:  
GENSCAN (http://CCR-08i.mit.edu/GENSCAN.html), GENESH  
(http://www.sotberry.com/), GeneMark.hmm  
(http://opal.biology.gatech.edu/GeneMark/), GlimmerM  
(http://www.tigr.org/cdb/glimmerm/glmr form.html), RiceHMM  
(http://rgp.dna.affrc.go.jp/RiceHMM/) "SplicePredictor  
(http://bioinformatics.iastate.edu/cgi-bin/sp.cgi), sim4  
(http://globin.cse.psu.edu/html/docs/sim4.html), gap2  
(http://www.tigr.org/software/glimmer/), BLASTN and BLASTX. The  
genomic sequence was searched against NCBI NonRedundant Protein  
database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA  
sequence database at RGP or DDBJ. Protein homologues of the coding  
regions were searched against NCBI NonRedundant Protein database  
with BLASTP. ESTs represent the identified cDNA sequences using  
BLASTN with the corresponding DDBJ accession no. and RGP clone ID.  
BLASTN with the corresponding DDBJ accession no.  
A gene with identity or significant homology to a protein is  
classified based on the protein name to indicate the homology level  
such as same name, 'putative-' and '-like protein'. A gene without  
significant homology to any protein but with full-length cDNA or  
EST homology (covering almost the entire length of partial  
sequence) is classified as an 'unknown' protein. A gene predicted  
by two or more gene prediction programs is classified as a  
'hypothetical' protein according to IRGSP standard. A gene  
predicted by a single gene prediction program is also classified as  
a probable 'hypothetical' protein and is included as a  
miscellaneous feature of the sequence.  
The orientation of the sequence is from T7 to SP6 of the PAC clone.  
This sequence of P0693E08 clone has an overlap with P0006C08 (DDBJ:  
AF004683) clone at 5' end and with OSUNBa003K18 (DDBJ: AF005875)  
clone at 3' end. Detailed information on overlap and assembly  
quality together with annotation of this entry is available at  
http://rgp.dna.affrc.go.jp/GenomeSeq.html.

FEATURES  
source  
1..157969  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="genomic DNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/chromosome="2"  
/clone="P0693E08"  
complement(1878..2063)  
/gene="P0693E08.1"  
complement(<1878..>2063)  
/notes="start and end point are not identified"  
complement(1878..2063)  
/gene="P0693E08.1"  
/notes="predicted by GlimmerM etc."  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="BAD28618.1"  
/db\_xref="GI:50251479"  
/translation="MMQWSCRRLLPVVCGIGDLTAGTGAGRRGRKRPTAYLPNKY  
RLRVEYTRPTNGRW"  
4350..5856  
/gene="P0693E08.2"  
complement(4350..5856)  
/gene="P0693E08.2"  
/notes="supported by full-length cDNA(s) : AK106502"  
4350..5856  
/gene="P0693E08.2"  
/notes="contains full-length cDNA(s) : AK106502  
non-coding transcript  
probably inactive due to including stop codon(s) in CDS"  
7496..13886  
/gene="P0693E08.3"  
join(7496..8090,11440..11508,11600..11668,12554..12640,

13245..13312,13410..13886)  
/gene="P0693E08.3"  
/note="supported by full-length cDNA(s) : AK067325"  
join(7629..8090,11440..11508,11600..11668,12554..12640,  
13245..13312,13410..13551)  
/gene="P0693E08.3"  
/note="contains full-length cDNA(s) : AK067325"  
/codon\_start=1  
/product="putative signal peptidase I"  
/protein\_id="BAD28619.1"  
/db\_xref="GI:50251480"  
/translation="MSSIHFFPATQAAAPSRRLTKPSKPPAATAAATHRGDNEPV  
LRFILFLRRRARPAAPAPVPGDSSSGGGGGGGGGDEEGTRKGLLP  
EKLSTVDDAKTVLAAIAISLAFSPFAEPIPSLSMFPTFDVGDRIVAEKVTYFR  
KPCINDIVIFKSPVLQEVGYTDNDVFIKRIVARGDVVEVHKGLVAVGEVNEEF  
LPPSYDMNPVOVPENSVFVMDGNRNSVDSHWGPLPSKNILGRSIFRYPWPGRIGS  
TTTDCILPEPTNPSLLIDVKLAK"  
15268..16496  
/gene="P0693E08.4"  
join(<15268..15743,16184..>16496)  
/gene="P0693E08.4"  
/note="start and end point are not identified"  
join(15268..15743,16184..16496)  
/gene="P0693E08.4"  
/codon\_start=1  
/product="alpha-expansin OsEXPA13"  
/protein\_id="BAD28620.1"  
/db\_xref="GI:50251481"  
/translation="MAGVARMLAUVCAIMPAAAMAGVGALPESGWRAHATFYGG  
ADAAGTMGACGYGLNYAGYGTAAALSTALFNDGLACGCYKLVCDRTKRTWCKP  
GVSVTITATNFCFNPMDLPSDGGNCNPPRPHFMDQPAWEKIGIYRGGIIPIYQRV  
PMWKGKGVFRTHDHYFOLVLLTNVGAAGSIKAMDVKGSKSPDMAMAHNMGAWHS  
LAYLTGQGLSRVTTTDTGQTLVFPNVVRPGWRFQGTFSNIOFK"  
complement(join(17966..18049,18114..18212))  
/gene="P0693E08.5"  
complement(join(17966..18049,18114..18212))  
/gene="P0693E08.5"  
/note="hypothetical ORF  
predicted by GENSCAN  
this category is not included in IRGSP standard"  
18903..21603  
/gene="P0693E08.6"  
join(<18903..19081,21594..>21603)  
/gene="P0693E08.6"  
/note="start and end point are not identified"  
join(18903..19081,21594..21603)  
/gene="P0693E08.6"  
/note="predicted by GENSCAN etc."  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="BAD28621.1"  
/db\_xref="GI:50251482"  
/translation="MLPAEWNDKRCWTFPAEPDINAKKKKKKKKKKKKKKKLQQA  
GNAETCGSTLSIYPRRD"  
complement(22195..23379)  
/gene="P0693E08.7"  
complement(22195..23379)  
/gene="P0693E08.7"  
/note="probably inactive due to 5' exon missing in CDS  
pseudogene, putative polypeptide"  
25993..27165  
/gene="P0693E08.8"  
complement(<25993..>27165)  
/gene="P0693E08.8"  
/note="start and end point are not identified"  
25993..27165  
/gene="P0693E08.8"  
/note="predicted by GENESH etc."  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="BAD28622.1"  
/db\_xref="GI:50251483"

```
/translation="MTLISDLPELPLPCTIAGHLHAAVVVVRHFAVCWDRDLRLYLPR
RPSFLPWLPAWPTDDDTAAAGACRCVFSRTTYTHAGGLGRDKRVAHYDYGASWV
GGLFVNPLTGRAAAACVDDPYLSDWIDNEGSRICIFSGDGLTLTCCFVADGPPLSIXGA
IWPDYQVWKRLGDDSDERACAVAVHDGAACVVDLARCVCYVHEIGPAAAGDQDTFLP
LPDPGKVRRSYLLELRLGELLASVLQDAGCTDDDDDDDDRLSVSHAFDIAVALN
ALDQDAVAGAGDPPSPSVKMDGATGDHVLFLGYPRSPFAVEAARYGGRVPGGS
AYFVGRSKPCRVYRCSEDDGTAATLVDTLPAWMDRCWMLPEPDIAPIVIGART"
29219. .31638
/gene="P0693E08.9"
join(<29219. .29733,29771. .30051,30102. .30302,30471. .30527,
31173. .31237,31417. .>31638)
/gene="P0693E08.9"
/note="start and end point are not identified"
join(29219. .29733,29771. .30051,30102. .30302,30471. .30527,
31173. .31237,31417. .31638)
/gene="P0693E08.9"
/note="predicted by FGESH etc."
/codon_start=1
/product="hypothetical protein"
/protein_id="BAD28623.1"
/db_xref="GI:50251484"
/translation="MTSVSLPELVCLIGRRLHTAINVVRHFAVCWDRQSLRHIPP
SPPPAALLPWLPAWPTDDDTAAAGACRCVFSRTTYTHAGGLGRDKRVAHYDYGASWV
WFINDKLVNPLTGRAAAACVDDPYLSDWIDNEGSRICIFSGDGLTLTCCFVADGPPLSIXGA
RLSTPTTAAAPPATVLAHPMLPDEPKVRHRSYLLLELRLGELLASVLQDQDDDD
RLSVSHAFDIAVALDPAAPPPVAAQRWRWRCPPDGTTRREEQEEAARA
YQORRQOQDLRIHVSDDLSPQVDSRLRMYSEHVKVVRAGRVRQAQGGGLGW
LAVGCNLI PVGQILLPTFRNHYEDGIDEDSKVTGI FEEDLPCELLPMPDPSWERPL
KPHLRPSPPTPPASDAVNAATDTAGSIFACCHYRLSRSGKGNERRRG"
32993. .33890
/gene="P0693E08.10"
join(<32993. .33468,33578. .>33890)
/gene="P0693E08.10"

Query Match 6.4%; Score 71.2; DB 8; Length 157969;
Best Local Similarity 56.7%; Pred. No. 2.2e-10;
Matches 194; Conservative 0; Mismatches 138; Indels 10; Gaps 3;

Qy 348 TTAGAGCAAGTATTAAGTCTAGTCAGCTGGCTATAAGATGTTCCACATCAACAAATC 407
Db 102782 TTAGAGCAGGATTATAGATCTTAGTCAGCGCGCTAGCATGTTACCCGTCAGCGAAT 102841
Qy 408 CTTAAACTGGAGGAGAAAGAAAGTAGGAGTGAGAGGCGCTCGCGCTTGTCTCAATCGCT 467
Db 102842 CCGTAG-TGGAGGATTGAGAGGGAAGAAATAGAAATGAGCGCGCTTCCATCTATCAC- 102899
Qy 468 AGCATAGCAAGCTCCATGGAATCGAGCCACATGCAACCGCCACATGACTAAGG 527
Db 102900 -CCGGCTGAAGCGCAACACATGAAGAAAATAATGATTTCTTCCAATCAGATCGAGAG 102958
Qy 528 CAAGCGCCAGCAATCAGTATGCTTTCTCTGCACTTTCTTCAAGCAATTAATAAC 587
Db 102959 ATCAGGTGCAATGTGAATATAAT-----ATCTAATCTTCGATAGATATAACAA 103011
Qy 588 TATAGCTAATCTAGCCAGTTTATATATAAAGAGGCTATATAGCTGACCTGGCAGTGC 647
Db 103012 ACAGCAACCTAAAGCCAGCTATTATATAGTATGATGAGTATAGTGACATGCTATTAG 103071
Qy 648 TATAGCCGCGCAGCGGCTCTTCTATTAGCTTTGCTCTTAT 689
Db 103072 TATAGCCGCGCAGCGGCTATGTTATTAACTGCTCTTAT 103113

RESULT 6
TAE404845 Triticum aestivum mRNA for thioredoxin h. linear PLN 15-AUG-2001
DEFINITION
ACCESSION AJ404845
VERSION AJ404845.1 GI:8980490
KEYWORDS thioredoxin h.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
```

```
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1. .629
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/tissue_type="aleurone"
/dev_stage="germinating seed"
/country="Spain"
1. .111
112. .489
/note="ORF"
/codon_start=1
/product="thioredoxin h"
/protein_id="CAB96331.1"
/db_xref="GI:8980491"
/db_xref="GOA:Q9LDX4"
/db_xref="UniProt/TREMBL:Q9LDX4"
/translation="MAASAATATAAAGAGEVISHLSQWTTMOIEANAAKLVVID
FTASWGPCRIMAPIADLAKKEPAVFLKVDVDELKSLAEQFSVEAMPTFLPMKEGD
VKDRVVGAIKBELTNKVGHLAAQ"
490. .629
polya_signal
516. .523

Query Match 6.2%; Score 68.6; DB 8; Length 629;
Best Local Similarity 91.4%; Pred. No. 5.4e-10;
Matches 85; Conservative 0; Mismatches 4; Indels 4; Gaps 1;

Qy 1019 GCGGCGCGGATTTGGAGACAGAGCCACAGGCAACAAAGTCGCGTGAGAAATCAA 1078
Db 1 GCGAGCGGATTTGGAGACCGAGCCAGAGGCAACAAAGTCGCGTGAGAAAT--- 57
Qy 1079 CAAGCGGTGTTTCCGAGAGAGAGAGAGAG 1111
Db 58 -AAGCGGTGTTTCCGAGAGAGAGAGAGAGAG 89

RESULT 7
AP005524
LOCUS
DEFINITION
Orzya sativa (japonica cultivar-group) genomic DNA, chromosome 8,
PAC clone:P0562A06.
ACCESSION
AP005524
VERSION
AP005524.3 GI:42409193
KEYWORDS
SOURCE
ORGANISM
Orzya sativa (japonica cultivar-group)
Orzya sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1
Sasaki,T., Matsumoto,T. and Katayose,Y.
Orzya sativa nipponbare(GA3) genomic DNA, chromosome 8, PAC
clone:P0562A06
Published Only in Database (2002)
JOURNAL
REFERENCE
2 (bases 1 to 154198)
```

```
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poodeae; Triticeae; Triticum.
1
Serrato,A.J., Crespo,J.L., Florencio,F.J. and Cejudo,F.J.
Characterization of two thioredoxins h with predominant
localization in the nucleus of aleurone and scutellum cells of
germinating wheat seeds
Plant Mol. Biol. 46 (3), 361-371 (2001)
21380673
11488482
2 (bases 1 to 629)
Cejudo,F.J.
Direct Submission
Submitted (06-JUL-2000) Cejudo F.J., Instituto de Bioquímica
Vegetal y Fotosíntesis, Universidad de Sevilla, Avda Americo
Vespucio s/n, 41092-Sevilla, SPAIN
Location/Qualifiers
```

FEATURES

source



**AUTHORS** Sasaki, T., Matsumoto, T. and Katayose, Y.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (10-JUL-2002) Takuji Sasaki, National Institute of  
 Agricultural Sciences, Rice Genome Research Program; Kannondai  
 2-1-2, Teukuba, Ibaraki 305-8602, Japan  
 (E-mail: tsasaki@nias.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/,  
 Tel: 81-298-38-7441, Fax: 81-298-38-7468)  
**COMMENT** On Feb 4, 2004 this sequence version replaced gi:38678125.  
 Genes were predicted from the integrated results of the following:  
 GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH  
 (http://www.softberry.com/), GeneMark.hmm  
 (http://opal.biology.gatech.edu/GeneMark/), GlimmerM  
 (http://www.tigr.org/cdb/glimmerm/glmr\_form.html), RiceHMM  
 (http://rgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor  
 (http://bioinformatics.iastate.edu/cgi-bin/sp.cgi), sim4  
 (http://globin.cse.psu.edu/html/docs/sim4.html), gap2  
 (http://www.tigr.org/software/glimmer/), BLASTN and BLASTX. The  
 genomic sequence was searched against NCBI Nonredundant Protein  
 database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA  
 sequence database at RGP or DDBJ. Protein homologies of the coding  
 regions were searched against NCBI Nonredundant Protein database  
 with BLASTP. ESTs represent the identified cDNA sequences using  
 BLASTN with the corresponding DDBJ accession no. and RGP clone ID.  
 Full-length cDNAs representing the identified cDNA sequences using  
 BLASTN with the corresponding DDBJ accession no.  
 A gene with identity or significant homology to a protein is  
 classified based on the protein name to indicate the homology level  
 such as same name, 'putative-' and '-like protein'. A gene without  
 significant homology to any protein but with full-length cDNA or  
 EST homology (covering almost the entire length of partial  
 sequence) is classified as an 'unknown' protein. A gene predicted  
 by two or more gene prediction programs is classified as a  
 'hypothetical' protein according to IRGSP standard. A gene  
 predicted by a single gene prediction program is also classified as  
 a probable 'hypothetical' protein and is included as a  
 miscellaneous feature of the sequence.  
 The orientation of the sequence is from T7 to SP6 of the PAC clone.  
 This sequence of P0562A06 clone has an overlap with OJ1150 All  
 (DDBJ: AP003928) clone at 5' end and with P0604E01 (DDBJ: AP005544)  
 clone at 3' end. Detailed information on overlap and assembly  
 quality together with annotation of this entry is available at  
 http://rgp.dna.affrc.go.jp/GenomeSeq.html.

**FEATURES**  
 source  
 1. 154198  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="genomic DNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /chromosome="8"  
 /clone="P0562A06"  
 /complement(2126..4572)  
 /gene="P0562A06.1"  
 complement(join(2126..2526,2641..2659,2763..2938,  
 3417..3463,4103..4572))  
 /note="supported by full-length cDNA (s): AK065151"  
 complement(join(2133..2526,2641..2659,2763..2938,  
 3417..3463,4103..4572))  
 /gene="P0562A06.1"  
 /note="supported by full-length cDNA (s): AK099458"  
 complement(join(2422..2526,2641..2659,2763..2938,  
 3417..3463,4103..4517))  
 /gene="P0562A06.1"  
 /note="contains EST (s): AU068303 (C12982), AU068304 (C12982)  
 contains full-length cDNA (s): AK065151, AK099458"  
 /codon\_start=1  
 /product="putative fiber protein Fb1"  
 /protein\_id="BAD13126.1"  
 /db\_xref="GI:45736095"  
 /translation="MADYNYRGYGGVSTPSAPPASSYGYTTTTPSAPSSSSSYCYG  
 HGGGYPSSTYPPPPSSQAYPMGGGLVFPFGTHPDVFRFAVDVDRGSSIDSR  
 ELQDALSAYHRSIRTVLLFLFNKPASHSPSRMGPAEFLVMNCLGQWRGIFDRI  
 DRDGSKIEKDREALSLGYAVPPSVLELLIANYNGVSSRGALDFDNFVBCGMIV  
 KGLTEKPKKEDTRYSGSATLSYDGLSMVPIFIVP"  
 /join(5570..5640,6229..6318,6443..6557)  
 /gene="P0562A06.2"  
 join(5570..5640,6229..6318,6443..6557)  
 /gene="P0562A06.2"  
 /note="hypothetical ORF  
 predicted by GENSCAN  
 this category is not included in IRGSP standard"  
 join(6892..6993,7148..7184,7470..7579)  
 /gene="P0562A06.3"  
 join(6892..6993,7148..7184,7470..7579)  
 /gene="P0562A06.3"  
 /note="hypothetical ORF  
 predicted by GENSCAN  
 this category is not included in IRGSP standard"  
 8510..11409  
 /gene="P0562A06.4"  
 join(8510..8859,8978..9042,9255..9333,9484..9502,  
 9583..9702,10049..10118,10201..10244,10410..10451,  
 10525..10584,10680..10747,10905..10962,11040..11409)  
 /gene="P0562A06.4"  
 /note="supported by full-length cDNA (s): AK069761"  
 join(8549..8859,8978..9042,9255..9333,9484..9502,  
 9583..9702,10049..10118,10201..10244,10410..10451,  
 10525..10584,10680..10747,10905..10962,11040..11129)  
 /gene="P0562A06.4"  
 /note="contains EST (s): AU173424 (R3033), AU173425 (R3033)  
 contains full-length cDNA (s): AK069761"  
 /codon\_start=1  
 /product="putative auxin-regulated protein"  
 /protein\_id="BAD13127.1"  
 /db\_xref="GI:45736096"  
 /translation="MAATLHLPPLLARLRFSSAAASTSTSRRTLSAQLDDTAA  
 ASTSTSTSPKPAASSFAPPKPEKTPFVKQSQSDDIVTASLAIPERLGTGVA  
 LGYSLVSPDEVADEYALDQGRKVKSSKIGQCPREKPEIIEYFEGCCPCKVR  
 ENVAVDLDFVLPCKNGPTFRPKVLEMGKQFPYMDPNTGMVAYESDAILKYLA  
 DKYDGTGVPLMSLGLTTITAGLAMSRSRSGSKYTPAKLPPELWYBGSPECK  
 IVRETLVELELPHLLHSCARGSPRQREFLKYGIFQAPYIEDPNTGKMFESADIIDY  
 LRATVAA"  
 12290..12868  
 /gene="P0562A06.5"  
 <12290..>12868  
 /gene="P0562A06.5"  
 /note="start and end point are not identified"  
 12290..12868  
 /gene="P0562A06.5"  
 /note="similar to Arabidopsis thaliana chromosome 3,  
 NP\_190024"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /protein\_id="BAD13128.1"  
 /db\_xref="GI:45736097"  
 /translation="MDSSPSAPASGRWSGVSALVGAASATAAAAVLCLRPDRPTFE  
 LISLSLSTFHFRPPAALDGLTIVHATNPVVRYPVYGPSTVSYLDGAHLGLTARLDA  
 GEQPTSCRLHLPARLDAVELAHARSILADTARRHMELDAAVKTAGAVALWSSR  
 FSVSDSHIVDPVFLDVEIQENHSEMQVLT"  
 complement(12959..19068)  
 /gene="P0562A06.6"  
 complement(12959..19068)  
 /gene="P0562A06.6"  
 /note="contains full-length cDNA (s): AK071334, AK060581  
 non-coding transcript  
 probably inactive due to including stop codon (s) in CDS"  
 complement(22111..25642)  
 /gene="P0562A06.7"  
 complement(22111..25642)  
 /gene="P0562A06.7"  
 /note="probably inactive due to 5' exon missing in CDS  
 pseudogene, RNA-directed DNA polymerase"  
 /pseudocd  
 complement(26564..29230)  
 /gene="P0562A06.8"  
 complement(join(26564..26981,27068..27130,27982..28102,

```
28202. .28403,28511. .28759,29097. .29230))
/ gene="P0562A06.8"
/ note="supported by full-length cDNA (s): AK066924"
complement(join(26584. .26981,27068. .27130,27982. .28102,
28202. .28403,28511. .28759,29097. .29187))
/ gene="P0562A06.8"
/ note="supported by full-length cDNA (s): AK104800"
complement(join(26838. .26981,27068. .27130,27982. .28102,
28202. .28403,28511. .28706))
/ gene="P0562A06.8"
/ notes="contains EST(s): AU100788(C50778),C97037(C50778)
contains full-length cDNA(s): AK066924,AK104800"
/codon_start=1
/product="putative vesicle-associated membrane protein 725
(AtVAMP725)"
/protein_id="BAD13129.1"
/db_xref="GI:45736098"
/translation="MNGKQSLIYSFVAKGSVLAETAFSGNFSTIAVCLQKLPN
TSKSTYSGDHTFNLVDRGFVLADEAVGRSPVFLDRVKEDFMQRYSIDEE
GQPLADADDDDFLLDRFSIAVLDREFGRLKDHMLYCIINHPEISKLKVAHL
TEVKGIMMDIKELRGEKIELLVGKTETLQSQADSFRHGRRLRKLQNLRFKL
MVGNAVALILFWLIIICGFKC"
complement(29750. .33273)
/ gene="P0562A06.9"
complement(join(29750. .30219,30364. .30625,31510. .31599,
33145. .33273))
/ gene="P0562A06.9"
/ note="supported by full-length cDNA (s): AK103638"
complement(join(29910. .30219,30364. .30625,31510. .31599,
33145. .33187))
/ gene="P0562A06.9"
/ notes="contains EST(s): AU223334(S6006),AU223335(S6006)
contains full-length cDNA(s): AK103638
similar to Arabidopsis thaliana chromosome 3, NP_566962"
/codon_start=1
/product="unknown protein"
/protein_id="BAD13130.1"
Query Match 5.9%; Score 65.2; DB 8; Length 154198;
Best Local Similarity 54.5%; Pred. No. 1.9e-08;
Matches 206; Conservative 0; Mismatches 153; Indels 19; Gaps 3;
Qy 346 GTTTAGGCAAGTATAATAGTCTTAGTCAGTCGGCTATAGATGTTCCACATCAGCAAA 405
Db 119628 GCTAAGAGCAAGTATAGTGAAGCAGTCAGCGCGGAAATAACTTACCAGCTCACCAA 119687
Qy 406 TCCTTAACTGGAGAGAAAGTAGGAGTGAAGAGGGCGTCCGGCTTCGTCATCG 465
Db 119688 ATCTGAGGTGAAAGAGAGAGAAATCCTAATAGAGAAATAGGACGGCGATTGGTGAGCG 119747
Qy 466 CTAGCGATGACCAAGCTCCCATGGAATCGAGGCAACATGCAACCCGACAAATGACTAAA 525
Db 119748 CTCGCTTGAAG-----CAGCAGAACCGAGATATAAAGTCTTCTCAGCCTGCTGCA 119799
Qy 526 GGCAAAACGCCAGCAATCAGTATGCTTCTTCATCTTTCTTCATGCAAGCATTAAT 585
Db 119800 TGGCACTGAGCGCTGAGCTCTCTTTCTCCCATTTGCTACTAGTACATGCAAGCATTAAT 119859
Qy 586 ACTATAGTAATCTACA-GCCAGTTTATATATAAACAAGGCTATAT-----AGCT 634
Db 119860 CTTATAGCAACCTCTACGCCAGTTTACTATATGCACTGGCTTTCTCAATGCTTAAAGT 119919
Qy 635 GACCTGGCAGTGCTATAGAGCGGCGAGCGGCTCTTCTATTAGCTTTGCTCTTATGGCTA 694
Db 119920 GACATGGCCACTTATTGGCCAGCTGCCGTGAACCATTAACCATGCTCTAATCGTCA 119979
Qy 695 CATCTGTGTGAGAGTCG 712
Db 119980 GATCAATCTCTGATGTCG 119997
RESULT 8
AP005544
```

```
LOCUS
DEFINITION
  Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 8,
  PAC clone:P0604E01.
ACCESSION
  AP005544
VERSION
  AP005544.3 GI:42409221
KEYWORDS
  Oryza sativa (japonica cultivar-group)
SOURCE
  Oryza sativa (japonica cultivar-group)
  Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
  Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE
  1
  Sasaki,T., Matsumoto,T. and Katayose,Y.
  Oryza sativa nipponbare(GA3) genomic DNA, chromosome 8, PAC
  clone:P0604E01
  Published only in Database (2002)
  2 (bases 1 to 159049)
  Sasaki,T., Matsumoto,T. and Katayose,Y.
  Direct Submission
  Submitted (18-JUL-2002) Takuji Sasaki, National Institute of
  Agrobiological Sciences, Rice Genome Research Program; Kannondai
  2-1-2, Tsukuba, Ibaraki 305-8602, Japan
  (E-mail:tsasak@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
  Tel:81-298-38-7441, Fax:81-298-38-7468)
  On Feb 4, 2004 this sequence version replaced gi:34740254.
  Genes were predicted from the integrated results of the following:
  GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH
  (http://www.softberry.com/), GeneMark hmm
  (http://opal.biology.gatech.edu/GeneMark/), GlimmerM
  (http://www.tigr.org/tdb/glimmer/glimr_form.html), RiceHMM
  (http://rgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor
  (http://bioinformatics.iastate.edu/cgi-bin/sp.cgi), sim4
  (http://globin.cse.psu.edu/html/docs/sim4.html), gap2
  (http://www.tigr.org/software/glimmer/), BLASTN and BLASTX. The
  genomic sequence was searched against NCBI NonRedundant Protein
  sequence database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA
  sequence database at RGP or DDBJ. Protein homologies of the coding
  regions were searched against NCBI NonRedundant Protein database
  with BLASTP. ESTs represent the identified cDNA sequences using
  BLASTN with the corresponding DDBJ accession no. and RGP clone ID.
  Full-length cDNAs represent the identified cDNA sequences using
  BLASTN with the corresponding DDBJ accession no.
  A gene with identity or significant homology to a protein is
  classified based on the protein name to indicate the homology level
  such as same name, 'putative-' and '-like protein'. A gene without
  significant homology to any protein but with full-length cDNA or
  EST homology (covering almost the entire length of partial
  sequence) is classified as an 'unknown' protein. A gene predicted
  by two or more gene prediction programs is classified as a
  'hypothetical' protein according to IRGSP standard. A gene
  predicted by a single gene prediction program is also classified as
  a probable 'hypothetical' protein and is included as a
  miscellaneous feature of the sequence.
  The orientation of the sequence is from SP6 to T7 of the PAC clone.
  This sequence of P0604E01 clone has an overlap with P0562A06 (DDBJ:
  AP005524) clone at 5' end and with P0543D10 (DDBJ: AP004587) clone
  at 3' end. Detailed information on overlap and assembly quality
  together with annotation of this entry is available at
  http://rgp.dna.affrc.go.jp/genomeSeq.html.
FEATURES
  Location/Qualifiers
    1. 159049
      /organism="Oryza sativa (japonica cultivar-group)"
      /mol_type="genomic DNA"
      /cultivar="Nipponbare"
      /db_xref="taxon:39947"
      /chromosome="8"
      /clone="P0604E01"
      47. .821
      /gene="P0604E01.1"
      join(<47. .148,395. .474,569. .>821)
      /note="start and end point are not identified"
      join(47. .148,395. .474,569. .821)
      CDS
      gene
      mRNA
```







Query Match	5.7%;	Score 62.8;	DB 2;	Length 169030;
Best Local Similarity	53.5%;	Pred. No. 1.1e-07;		
Matches 183;	Conservative	0;	Mismatches 147;	Indels 12; Gaps 2;
QY	347	TTTAGAGCAAGTAAATAAGTCTCTAGTCAGCTGGCTATAGATGTTCCACATCAGCAAAAT	406	
DB	77211	TTAAGAGCAGGTTTATATACATCTTAATCGCGCGCTAGCAGCTACGCGTCAGCCAA-	77153	
QY	407	CCTTAAACTGGAGGAGAAAGAAAGTAGGAGTGAGAGGGCGTCGGCGTTCGTCGAATCGC	466	
DB	77152	ATTTCAGCTGGAGGAGTGAGAGAGGGAAGAAAAGGATAGAGCGGCGGCTTCCATCTATCGC	77093	
QY	467	TAGCGATGACACAGCTCCCATCGAATCGAGCCAAACATGCAACCCGACCAATGACTAAAG	526	
DB	77092	CCGGCTGAAGCCGACACACAGGAGAAAAGCTGCTTCTTCCAAACAGATGCGAGGAAATC	77033	
QY	527	GCAACGCCAGCAATCAGTATGCGCTTCTCTGCAATCTTCTTCATGCAAGCATTTAAATA	586	
DB	77032	AGGTGTACATGCGAATATTAAAT-----ATCTAAATCTTCAAAAGAGGTAACA	76984	
QY	587	CTATAGCTAATCTACAGCCAGTTTATTATATAACAGGCTATATAGCTGACCTGGCAGTG	646	
DB	76983	AATAGTCAACTAACAGCCAGTCTATTATATTATTGATGCTATAGGTGACATGGTACTA	76924	
QY	647	CTATAGAGCCGCGCAGCGGCTCTTCTATTAGCTTTGCTCTTA	688	
DB	76923	GTATAGAGCCGCGCAGGTGCTCTATATTATTATCTTCTGCTCTAA	76882	
RESULT 11				
AP006149				
LOCUS	AP006149	171257 bp	DNA	linear
DEFINITION	Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 9, BAC clone: B1274F11.			
ACCESSION	AP006149			
VERSION	AP006149.2	GI:41152748		
KEYWORDS				
SOURCE	Oryza sativa (japonica cultivar-group)			
ORGANISM	Oryza sativa (japonica cultivar-group)			
REFERENCE	Sasaki, T., Matsumoto, T. and Katayose, Y.			
AUTHORS	Oryza sativa nipponbare (GA3) genomic DNA, chromosome 9, BAC clone: B1274F11			
TITLE	Published Only in Database (2002)			
JOURNAL	2 (bases 1 to 171257)			
REFERENCE	Sasaki, T., Matsumoto, T. and Katayose, Y.			
AUTHORS	Direct Submission			
TITLE	Submitted (26-DEC-2002) Takuji Sasaki, National Institute of Agrobiological Sciences, Rice Genome Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan			
JOURNAL	(E-mail: tsasakion@affrc.go.jp, URL: http://irgp.dna.affrc.go.jp/, Tel: 81-298-38-7441, Fax: 81-298-38-7468)			
COMMENT	On Jan 23, 2004 this sequence version replaced gi:27374998. Genes were predicted from the integrated results of the following: GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH (http://www.softberry.com/), GeneMark.hmm (http://opal.biology.gatech.edu/GeneMark/), GlimmerM (http://www.tigr.org/tdb/glimmer/glmr.form.html), RiceHMM (http://irgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor (http://bioinformatics.iastate.edu/cgi-bin/sp.cgi), sim4 (http://globin.cse.psu.edu/html/docs/sim4.html), gap2 (http://www.tigr.org/software/glimmer/), BLASTN and BLASTX. The genomic sequence was searched against NCBI NonRedundant Protein database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at RGP or DDBJ. Protein homologies of the coding regions were searched against NCBI NonRedundant Protein database with BLASTP. ESTs represent the identified cDNA sequences using BLASTN with the corresponding DDBJ accession no. and RGP clone ID. Full-length cDNAs represent the identified cDNA sequences using			

BLASTN with the corresponding DDBJ accession no.  
A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with full-length cDNA or EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted by two or more gene prediction programs is classified as a 'hypothetical' protein according to IRGSP standard. A gene predicted by a single gene prediction program is also classified as a probable 'hypothetical' protein and is included as a miscellaneous feature of the sequence.  
The orientation of the sequence is from -21M13 to M13rev of the BAC clone. This sequence of B1274F11 has an overlap with P0229B10 (DDBJ: AP006174) clone at 5' end and with P0478E02 (DDBJ: AB109206) clone at 3' end. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://irgsp.dna.affrc.go.jp/GenomesSeq.html>.

Location/Qualifiers

1. 171257

/organism="Oryza sativa (japonica cultivar-group)"

/mol\_type="genomic DNA"

/cultivar="Nipponbare"

/db\_xref="taxon:39947"

/chromosome="9"

622..1658

/clone="B1274F11"

/gene="B1274F11.1"

join(<622..714,1056..1147,1382..>1658)

/gene="B1274F11.1"

/notes="start and end point are not identified"

join(622..714,1056..1147,1382..1658)

/gene="B1274F11.1"

/notes="similar to Gossypium hirsutum, AAN05792"

/codon\_start=1

/product="hypothetical protein"

/protein\_id="BAD46562.1"

/db\_xref="GI:52076049"

/translation="MGNGSSGILSHSTVEKLYAWKKLFLEKVSASVSMICAML LKPIEVKGSFQLYVHGRDPLVLSMTAQLAQSVYGLQKHDKKIGLLRQEVKGVLYL KMEKNMETESLDSKMLVATQSIETTTSIMRLRESELFPOLLEVLAVLQV"

2855..3583

/gene="B1274F11.2"

<2855..>3583

/gene="B1274F11.2"

/notes="start and end point are not identified"

2855..3583

/gene="B1274F11.2"

/note="similar to Gossypium hirsutum, AAN05792"

/codon\_start=1

/product="hypothetical protein"

/protein\_id="BAD46563.1"

/db\_xref="GI:52076050"

/translation="WVQQLYLNLSLTNPTSNVHRQAALQLEIEVDWYSAFCSLVK SRDYYSLTGLRLSLFOSYHDPYKAKHNSDIYSLCEWOLAIIDIPDKVASEGK TLTIVHAVVQQAQEQKKSESFKLEKAEKLSLESKYGYSAGYGVDSR KSEVSDKRAKEALRKADEKSKYKSGIVTKMTLNINLTQTFPNVFMQMTGFASVC MEAFESVYFNKSDRILDSKRLT"

complement(join(4036..4453,4559..4690,4807..5160,5250..5496,5578..5654,6060..6134,6207..6308,6418..6536,6950..7073,7170..7220,7336..7434,7928..8088,9479..9690,13290..13633))

/gene="B1274F11.3"

complement(join(4036..4453,4559..4690,4807..5160,5250..5496,5578..5654,6060..6134,6207..6308,6418..6536,6950..7073,7170..7220,7336..7434,7928..8088,9479..9690,13290..13633))

/note="supported by full-length cDNA(s): AK068006"

complement(join(4036..4453,4559..4690,4807..5160,5250..5496,5578..5654,6060..6134,6207..6308,6418..6536,6950..7073,7170..7220,7336..7434,7928..8088,9479..9690,13290..13633))

/note="supported by full-length cDNA(s): AK065681"

FEATURES	source
gene	
mRNA	
CDS	
gene	
mRNA	
CDS	
gene	
mRNA	
CDS	

```
CDS
complement(join(4325..4453,4559..4690,4807..5160,
5250..5496,5578..5654,6060..6134,6207..6308,6418..6536,
6950..7073,7170..7220,7336..7434,7928..8038))
/notes="B1274F11.3"
/gene="contains EST(s): AU101494(E60352),AU030858(E60352)
contains full-length cDNA(s): AK068006,AK065681,AK068206"
/codon_start=1
/product="putative Clb1 protein"
/protein_id="BAD46564.1"
/db_xref="GI:52076051"
/translations="MGVISTVLGSGRFGFSAGIVGVYFFIYFQPTDKOVKVRPL
VEYDSISLDGILPEILPVKNPDYRIDNLNRFLEMMWPFYLNKAICTETALDIAPPIE
ENRKYTSIESIESIFSLPTFGMKVIVTEQEELIMPELSKWAAPNVTVVVKA
YGLKATQIVDLQVFASPRITLKLPTPEFCFAKILVLSMEKPHVDFGLKFGADLMA
IPGLYFRVQTIKKQVASYLWPKTLEVPIMPSKASKKPGVIGILLVKVLRQNLKRD
LIGKSDPYVVKMSDDKLPKSTKTRKSNLNPENWEDFKPVVDPDTQALEINVFQWE
OVGKHEKMGWNNLLKELPADETKVMTNLLKTMDPNDVQNEKSGQLTLEVYKPEPK
EEDMEKEGINADVVEKAPDGTAGGGLLYVVVHEAQDLLEGKHTNPYAKILIPKGEK
DVIYIKRNPDRWEDEFVCEBPPVNDKLIHIVLSKASKKGLIHGKETLIGYIDISLA
DVISNRKINSYHLIDSNGQIQIEMQWRTS"
join(12777..12934,13027..13124,13212..13816)
/gene="B1274F11.4"
join(<12777..12934,13027..13124,13212..>13816)
/gene="B1274F11.4"
/notes="start and end point are not identified"
join(12777..12934,13027..13124,13212..13816)
/gene="B1274F11.4"
/notes="contains EST(s): C19154(E10033),AU078664(E10033)"
/codon_start=1
/product="unknown protein"
/protein_id="BAD46565.1"
/db_xref="GI:52076052"
/translations="MAATGDAAVGELRAGRGGTARGLTRQPPTRRQGEELQTG
CGEGVARGVQGRRCRRGATVARNKMQIQSDGRSIATDVREEVKCSYLCSRAAS
RTEKATGVPWASRRKDLAGSPESRRARAADAQTAERARAAAAVAASRSTAC
RKVSDRELTAVDQGPAPLMQVLDGIEKMLNDYMPAAAAVAGNGMAAPSSRUSA
DHVYLRKVRDQGEERKEREDADVATLTCGAHVDPDLSQLPHRIKPEKPLRASVMVD
VTTTSKFIPT"
15612..16154
/gene="B1274F11.5"
<15612..>16154
/gene="B1274F11.5"
/notes="start and end point are not identified"
15612..16154
/gene="B1274F11.5"
/notes="contains EST(s): D23552(C2994),C71832(E0426)"
/codon_start=1
/product="unknown protein"
/protein_id="BAD46566.1"
/db_xref="GI:52076053"
/translations="MENAPKLSISSAKSPSFMASNKHYHGVNHHQPPFPPTPPQ
PLPLPARHAPPQOHAAPAPQPSKPSHATATAAAAAALGPKKTAPVPTPVQAA
PSKRAAASQQAEEWTTTSDLYSVSPSPSVPMTSLLVGAAGRAKAAATACAVEV
AGGGVDVGGATDRLRLRL"
complement(16657..20028)
/gene="B1274F11.6"
complement(join(16657..17058,17236..17379,17480..17542,
18082..18205,18868..19094,19177..19344,19437..19587,
19656..20028))
/gene="B1274F11.6"
/notes="supported by full-length cDNA(s): AK071977"
complement(join(16657..17058,17236..17379,17480..17542,
18082..18205,18868..19094,19177..19344,19437..19587,
19656..20025))
/gene="B1274F11.6"
/notes="supported by full-length cDNA(s): AK098959"
complement(join(16927..17058,17236..17379,17480..17542,
18082..18205,18868..19094,19177..19344,19437..19587,
19656..19978))
/gene="B1274F11.6"
/notes="contains EST(s): AU056551(S20675),AU162699(S20675)
contains full-length cDNA(s): AK098959,AK071977"
/codon_start=1
```

```

/product="putative dehydroquinase synthase"
/protein_id="BAD46567.1"
/db_xref="GI:52076054"
/translations="MAAAASSSLLAAASSSSRAAAVSARRAPSAPAAASLPSPSRA
SCAPLRASAEATLSRVVASAPQPPASRVSTVDVLDGRSYPLVIGLGLDE
POLLQHVHGRKRVLVVNTITVAPFLDKETVWALTHTNNPNVSVESVILPOEKYKMGDT
LMKVDFKAVESLDRRCCTFVALGGVIGDMCGFAAAFLRGVNFIOIPTLLMAQVDSS
VGGKTGINHPLGNLIGAFYQPCVLIDTETLNTLPDLREASGIAEVVYGLIRDAPF
FEQBNKMPALLAREPSALAYAIKRSCENKAEVQAQDEKESGLRATNLNHTFPGHAIE
TGTGCAWLHGEAAGTVAADMASHRLMGIDESIKRAIDILEKAKLPITPEAMTV
EKFKSTMAVDKVKVADGLLRLLKGLPSGCVTGYDRNALDETLEAFCCD"
20888..226689
/gene="B1274F11.7"

Query Match      5.7%; Score 62.8; DB 8; Length 171257;
Best Local Similarity 53.5%; Pred. No. 1.1e-07;
Matches 183; Conservative 0; Mismatches 147; Indels 12; Gaps 2;

QY 347 TTTAGAGCAAGTATATAAGTCTCTAGTCAGCTGGCTTAAGATGTTTCCACATCAGCAAT 406
Db 97286 TTAAGAGCAGGTTTATATACATCTTAATCCGCGCGCTAGCACGCTACGCTCAGCCAA- 97344
|||
QY 407 CCTTAAACTGGAGGAGAAAGTAGGAGTGAGAGGGCGCTCGCGCTTCGTCAATCGC 466
|||
Db 97345 ATTGACGTGGAGGAGTGAGAGGGAAGAAAGGATAGAGCGGCGGTTCCTCTATCGC 97404
|||
QY 467 TAGCGATAGCAACAGCTCCCATCGAATCAGACCAACATCAACCCGCCACAAATGACTAAAG 526
|||
Db 97405 CCGGTGAAGCCACACACAGAGAAAAGCTGCTTCTTCCACACAGATGCGAGGAATC 97464
|||
QY 527 GCAACGCCAGCAATCAGTATGCCTTTCTCTGCACTTTTCTTCATGCAAGCAATTAATA 586
|||
Db 97465 AGGTGTACATCGCAATATTAAT-----ATCTAAATCTTCAAAGAGGTAACA 97513
|||
QY 587 CTATAGCTAATCTACAGCCAGTTTATATATAACAGGCTATATAGTACCTGCGCAGTG 646
|||
Db 97514 AATAGCTAACCTAACAGCCAGTCTATATATATTATGATGCTATAGTGACATGGTACTA 97573
|||
QY 647 CTATAGCGCGCAGCGCGCTCTTCTATTAGCTTTGCTCTTA 688
|||
Db 97574 GTATAGCGCGCAGCGTGGCTATATATTATCTTGCTCTAA 97615

RESULT 12
LOCUS AC079853/c 152423 bp DNA linear PLN 12-MAY-2001
DEFINITION Oryza sativa, complete sequence.
ACCESSION AC079853
VERSION AC079853.2 GI:13384340
KEYWORDS HTG.
SOURCE Oryza sativa
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 152423)
Wing,R.A., Frisch,D., Presting,G., Wood,T., Yu,Y., Soderlund,C.,
Kim,H.-R., Rambo,T., Henry,D. and Simmons,J.
Rice Genomic Sequence
Unpublished
2 (bases 1 to 152423)
Wing,R.A., Frisch,D., Presting,G., Wood,T., Yu,Y., Rambo,T.,
Henry,D. and Simmons,J.
Direct Submission
Submitted (14-SEP-2000) Clemson University Genomics Institute,
Clemson University, 100 Jordan Hall, Clemson, SC 29634, USA
3 (bases 1 to 152423)
Wing,R.A., Frisch,D., Presting,G., Wood,T., Yu,Y., Soderlund,C.,
Kim,H.-R., Rambo,T., Henry,D. and Simmons,J.
Direct Submission
Submitted (11-APR-2001) Clemson University Genomics Institute,
Clemson University, 100 Jordan Hall, Clemson, SC 29634, USA
4 (bases 1 to 152423)
```





[illegible]

TITLE	Vanaken,S.S., Riedmuller,S.B., Utterback,T.T., Feldblyum,T.V., Yang,Q.Q., Haas,B.J., Suh,B.B., Peterson,J.J., Quackenbush,J., White,O., Salzberg,S.L. and Fraser,C.M.
JOURNAL	Oryza sativa chromosome 3 BAC OSJNBa0087G11 genomic sequence
REFERENCE	2 (bases 1 to 169728)
AUTHORS	Buell,R.
TITLE	Direct Submission
JOURNAL	Submitted (06-JUN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
REFERENCE	3 (bases 1 to 169728)
AUTHORS	Buell,R.
TITLE	Direct Submission
JOURNAL	Submitted (08-MAY-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
REFERENCE	4 (bases 1 to 169728)
AUTHORS	Buell,R.
TITLE	Direct Submission
JOURNAL	Submitted (02-APR-2003) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
REFERENCE	5 (bases 1 to 169728)
AUTHORS	Buell,R.
TITLE	Direct Submission
JOURNAL	Submitted (17-APR-2003) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org
REFERENCE	6 (bases 1 to 169728)
AUTHORS	Buell,R.
TITLE	Direct Submission
JOURNAL	Submitted (23-APR-2003) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org
COMMENT	On Apr 2, 2003 this sequence version replaced gi:20503084. Address all correspondence to:rice@tigr.org
	BAC clone OSJNBa0087G11 is from Oryza sativa chromosome 3
	The orientation of the sequence is from SP6 to T7 end of the BAC clone.
	Genes were identified by a combination of several methods: Gene prediction programs including Fgenesh ( <a href="http://www.softberry.com/">http://www.softberry.com/</a> ), genscan and Genscan+ (Chris Burge, <a href="http://CCR-081.mit.edu/GENSCAN.html">http://CCR-081.mit.edu/GENSCAN.html</a> ), GeneMarkHMM (Mark Borodovsky, <a href="http://genemark.biology.gatech.edu/GeneMark/">http://genemark.biology.gatech.edu/GeneMark/</a> ), contact mpertea@tigr.org), (Mihaela Perlea and Steven Salzberg, <a href="http://www.tigr.org/tdb/tgi.shtm1">http://www.tigr.org/tdb/tgi.shtm1</a> ). The plant EST database at TIGR ( <a href="http://www.tigr.org/tdb/tgi.shtm1">http://www.tigr.org/tdb/tgi.shtm1</a> ). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown proteins'. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <a href="http://genome.wustl.edu/eddy/tRNAscan-SE/">http://genome.wustl.edu/eddy/tRNAscan-SE/</a> ). Simple repeats are identified by repeatmasker (Arian Smit, <a href="http://ftp.genome.washington.edu/RM/RepeatMasker.html">http://ftp.genome.washington.edu/RM/RepeatMasker.html</a> ).
FEATURES	This BAC overlaps with rice BAC OSJNBb004M10 (AC079853) and OSJNBa0015X03 (AC084295).
source	Location/Qualifiers 1. .169728 /organism="Oryza sativa (japonica cultivar-group)" /mol_type="genomic DNA" /cultivar="Nipponbare" /db_xref="taxon:39947" /chromosome="3" /map="near SL164" /clone="OSJNBa0087G11" /notes="japonica cultivar-group"
repeat_region	5. .92 /rpt_family="AT_rich"
repeat_region	1300. .1343 /rpt_family="AT_rich"
gene	complement(2402. .2977)

FEATURES	This BAC overlaps with rice BAC OSJNB00004M10 (AC079853) and OSJNBa0015X03 (AC084295).
source	Location/Qualifiers 1. .169728 /organism="Oryza sativa (japonica cultivar-group)" /mol_type="genomic DNA" /cultivar="Nipponbare" /db_xref="taxon:39947" /chromosome="3" /map="near S1764" /clone="OSJNBa0087G11" /note="japonica cultivar-group"
repeat_region	5..92 /rpt_family="AT_rich"
repeat_region	1300..1343 /rpt_family="AT_rich" complement(2402..2977)
gene	

```
/gene="OSUNBa0087G11.1"
/note="similar to 15 kDa organ specific salt induced
protein GI:AB23484 GI:256638 (Oryza sativa)"
complement(<2402..>.2977)
/gene="OSUNBa0087G11.1"
complement(2402..2977)
/gene="OSUNBa0087G11.1"
/codon_start=1
/product="putative salt-induced protein"
/protein_id="AAP12924.1"
/db_xref="GI:30017502"
/translation="MAAATNELQFVWVQLQAGSTVAKIGPWGGDYGGRRDHVTVAPRR
LRVSLRHGKIIDSIAFTYDGGDGLHSVPGWGGGAELPAVARKLAAGRRPFGA
TVABETPDAGERTVEHGTGPGFGDRSLVSLKLVDRRTIGPFGYAGTSPFVPR
GDGVGFFVRAGVLEAIGVYVNPICPSEK"
2672..2699
/rpt_family="GC_rich"
2840..2922
/rpt_family=" (CCG)n"
4017..4040
/rpt_family="AT_rich"
4278..4457
/rpt_family="T-rich"
5450..5485
/rpt_family="AT_rich"
6151..6172
/rpt_family="AT_rich"
12537..15658
/gene="OSUNBa0087G11.2"
/note="predicted by GenemarkHMM"
join(<12537..12600,13984..14478,15561..>.15658)
/gene="OSUNBa0087G11.2"
join(12537..12600,13984..14478,15561..15658)
/codon_start=1
/product="hypothetical protein"
/protein_id="AAP12942.1"
/db_xref="GI:30017520"
/translation="MACYLIGQTNKKIRKOKVLSGRRRGSAGLALRVELVLGGVGGD
LEADAQWRRRRGGARGAVNRRRRFVRDQWRRGGGAVRRRLIIVDGGAVWGR
RRVGGGABETGEALMTAVTICBGRLLVGGRGGEARGGVVGRWRIVGGSDGGA
RGGSGEPAMAPCGFRGWRCAVGAQSLLHLPLSRGQWVRASGQRAPRAQGP
"
14070..14399
/rpt_family=" (CCG)n"
14778..14812
/rpt_family="AT_rich"
17685..17688
/rpt_family="AT_rich"
18398..18420
/rpt_family="GC_rich"
19047..19069
/rpt_family="AT_rich"
19484..30211
/gene="OSUNBa0087G11.3"
/note="similar to phosphatidylinositol 3,5 kinase
GI:CAC42810 GI:14571648 (Candida albicans)"
join(<19484..20085,20498..21200,21386..21580,21660..>.22004,
23177..24127,24221..24552,24674..24797,24962..25117,
25966..27447,28360..28620,29906..>.30211)
/gene="OSUNBa0087G11.3"
join(19484..20085,20498..21200,21386..21580,21660..22004,
23177..24127,24221..24552,24674..24797,24962..25117,
25966..27447,28360..28620,29906..30211)
/gene="OSUNBa0087G11.3"
/codon_start=1
/product="putative phosphatidylinositol 3,5-kinase"
/protein_id="AAP12926.1"
/db_xref="GI:30017504"
/translation="MDASDKTFADVWKLTSWLPRRSNPNVSRDFWMPDHSRCVYCD
EDTOFTIFNRHRCRCRIFCGKCTTNSIPASSGPDRIIDEGDKIRVCNFCFKWEO
BRAANKOMPLLSPEASLFSFKSAITINSVSTAGSYSTGHYOHVARASSISP
KSSQDKVCHDMLDTHVPEKMSAVSNKDETSVHFHYTNRSDEESECYASDRQV
```

```

ORQHQHNDHYGPDDEFDEBESSYNPITISPTVEBNVISKEVSSHATDQGFSTPLVTKMD
DEPDNDNSECCEARRSSSIYALESTDTNPLDFEKNELFWFPEPEDEDEMEVGLFDDDD
DEPDVADSORRIARSSSPGSGEFNRDRSSSEHKVMKNVVDHGFALISOLLOVEN
ISLHEGETGWLIEIVTSVSWEAANFLRPDTSGGMDPGYVKFTYVLKASNCFPST
VVGVCCKNVAHRMTSRIEKPRLLLLAGALEYQRTVQLSSIDILLQOETHLKMA
VAKIQAQPNLLLVHTSVRYAQDILLERNISLVNLIKRPDLDRDIARCTNAHIVPSID
LLPSQKLHCHELYFYDKVYVHSVNSNATKMPKTMFPEGCPKPLGTVLLKGGSD
ELKKIKHVYVYGIFAAYHLALETSLADRGATLPEIPLESPLTVALPDRSTRADSSIS
TVPGFTFVNSRQRTDGTDEHPHVAGSIRSTDPGTVDPVPSNECTTQRTTTFSSHSGT
WSANGSLANSKTVDRIEKATATSAITSGVLMDSHYTSYLEKNWYSGDYHEYGVMSWD
VTMTVTLANSNGSCHGTSSEASTNITNFAULKPEPDGSDILANVENVTNSNVQVP
VSTAVQNETQGHSTNKEIMASDHQSILVALSIRCVWKGITICERSHMLRIKTY
GNFDPGLRFLRDLCLFNQGYCICISDKPPEAHVCHYTHQOGLTISIRKHTEFVLPE
NDPVPSSAIVPDCVASSSESEHVTDTPSYAVFLNKGDSADQSNLMDPLQYLR
ALNKQWRNRNFDALNEYTPVHVSFLRTVEROVGPFLPIGVNDTVVGIYDDDEPTSI
ISVALASHEVHLQSLDESDTDDNSLVTDLRGASLTESVDETAGSELRSFVSTEDN
ILYLSGKNPSPSDPLAYEKASHIKVNFGEGLPGOVKYTVICYAKQDFALARIICCP
SERDFVRLSRCKKGARGCKSNVFFAKSLDRPIIKQVTKTELESFPMKFPAPYFGYI
SESIVTGSPCTIAKILGIYQVKSGLGGKEMDVLVNMELLPFHHVIRLIDLKGSYSL
RYNPDNSNGSNKVLQDNLLEAMPTSPIFVGNKAKRLLERAVMNDTAPLASIGVMDYSL
LVNGVDEKHELVMGIIDFMROYTWDKHLTWTVKTSIGLGGPKNVAPTISPQYKMRF
RKAMSTYFLVVPDOWPPAVPSPKQGAENQDND"
31529..31599
/rpt_family="AT_rich"
31807..31829
/rpt_family="AT_rich"
33538..33557
/rpt_family=" (CCG)n"
35422..35607
/rpt_family="T-rich"
35734..35762
/rpt_family="AT_rich"
36143..36190
/rpt_family="AT_rich"

Query Match 5.6%; Score 62.6; DB 8; Length 169728;
Best Local Similarity 53.3%; Pred. No. 1.3e-07;
Matches 184; Conservative 0; Mismatches 149; Indels 12; Gaps 2;

QY 344 CCGTTTAGAGCAAGTATAAAGTCCTTAGTCAGCTGGCTATAAGATGTTCCACATCAGCA 403
Db 164555 CCCATGAGAGCAGGTTTAAATAGATCCTAATCAGCGCGCTAGCAGCTACGGCTCAGCC 164614

QY 404 AATCCTTAACTGAGGAGAGAAAGAAAGTAGGAGTGAGAGGGCGTCGGCGCTTCGTCAT 463
Db 164615 AAATCTGAC-GTGGAGGAGTGAGAGAGGAGAAAGGATGGAGCGGGCGCTCCATCTAT 164673

QY 464 CGCTAGCATAGCACCAAGCTCCCATCGAATCGAGCAACATGCAACCCGCAACATGACTA 523
Db 164674 CGCCGGCTGAGCGGCACACATGAGAAAAGCTGCTTTCTTCAACCATGCGAGGAG 164733

QY 524 AAGCGAAAACGCCCAACATCAGTATGCTTCTTCGATCTTCTTCGATCATTCTTCATGCAAGCATAA 583
Db 164734 ATGAGGTGTACATCGCAATATTAAAT-----ATCTAAATCTTCAATAGAGGTAA 164782

QY 584 ATACTATAGCTAATCTACAGCGGATTTATATATAAACAGGCTATATAGCTGACCTGGCA 643
Db 164783 ACAACAGCTTAACCTAACGCCGCTTATATATATAGTATAGTGTAGTGACATGGTA 164842

QY 644 GTGCTATAGAGCCGCGAGCGGCTCTTCTATTAGCTTTGCTCTTA 688
Db 164843 CTAGTATAGAGCCGCGAGCTGGCTATATATATCTTGTCTGA 164887

RESULT 14
AC121366 AC121366 127901 bp DNA linear PLN 02-APR-2004
LOCUS
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 5 clone
```



/translation="MALPLSSPLTSSIKLWTPSRARREPLPLAAGALEPANTPTTR  
HHYSRHYRGLEGKEDGVGVLEWRIRERENEGVRFPKVEGGRLDFSQ  
VDLLIGNEKIGVHLQKICISVYGFELDQKQVDFCRCDQL"  
36936..40019  
/gene="OSUNBb0111K12.6"  
/note="unknown protein"  
CDS  
join(36936..37333,37523..37833,37955..38017,38594..39078,  
39669..40019)  
/gene="OSUNBb0111K12.6"  
/note="similar to rice EST AK120909"  
/codon\_start=1  
/product="unknown protein"  
/protein\_id="AAS79734.1"  
/db\_xref="GI:46063431"  
/db\_xref="UniProt/TsEMBL:Q8LQ49"  
/translation="MAAAGMGNGGAGDADKARDVTDOSKALGNSCEDRALPSAVRV  
TVSGDPVTGSGFNADYVHLHOPDEGDHGDSTECSSFGPSCSADSDDDDDTK  
SEMDGMDVSPFFGPTGTGADRASSAPRVRRRQVTAWRKI VGPIMRCQWLELNMK  
NLJSQVAKYDRELAINHEKDQLQEMVKAQPKSEPGKLYPQSHERILIMRRKKRDE  
DTVTSYLRHPALSYENKSGVQTDGPLVNGFDSVVEDIESTDDALVENDRVF  
EQSLREILITVDVQSRILSLQSRNARSVKYKLSQCLDRKQVQPKIQKQMTCC  
KQDGRSHQTKCMHTLQKDDLSLAVVPVFGSTDCVLECKMKNDQAQDAVSD  
LNGITIMFGCKNFVNAHVGLYKESADDDVLDNQAQEGYQLPEKVPKPSHSLV  
MPPSKVQKASADIVDYQVETAPVAKYIISGDRGQPKKKGLPVLAKIKITEKDP  
GNMKKTVLIVADVPRSTRVRKPKTY"  
complement(43031..46939)  
/gene="OSUNBb0111K12.7"  
/note="putative disease resistance protein, contains  
NBS-LRR domain"  
CDS  
complement(join(43031..43218,43419..44252,44326..46939))  
/gene="OSUNBb0111K12.7"  
/note="similar to rice EST CF326176, AK120289"  
/codon\_start=1  
/product="putative disease resistance protein, contains  
NBS-LRR domain"  
/protein\_id="AAS79735.1"  
/db\_xref="GI:46063432"  
/db\_xref="dbEST:33800613"  
/db\_xref="UniProt/TsEMBL:Q8LQ88"  
/translation="MAEVALASRLAASPILNKLLADASTYLVGMASELRELETSTM  
POFELLIEAEKGNHRAKDKWIRELAELYNALDELDDEHYDILKHKVNGGEDPSP  
DLEHASIGSIILKPMRAASSLSNLRPNIKLVRLQKELKAILAKARDPREMLGLPA  
GSSVEGAQTGHKTVVTAATSPFPKVGFRDADRIRIVDLTQHKTCASAPFVSI  
VPGMGKSTLAQVYNDKTIQEHFDVTMVCISRKDDVHRHTREIIESATKBEKQRV  
GNMDVLYKLEIKEDVLLVDDFDKSDVEWDLAPILSSQNGATKVLVT  
SRKSTPLPFLSEDIIDLENNKDTFOALFKHAFSGATIRDLQCMGCFEHAVKITE  
RLGSPPLAAKVGSLNLRVNIIDWKGALTIKIDNLEPKRALLMSYQKLDPCLOKCF  
LYCSLFPKGYKIIDEVLHVAEGFIDARTNKRMDTGMDFYKENVSSFFQPSSE  
RFDSTVIMHDLHLAESLSRDECDRLEDDKVRKIPCTVRHLVSURVESIIQHKPSVC  
KLQHLRTLICIDPLVDVGSNI FEQVNLNKLQVLYLSFYNTKPLPESIQGLKHLRYL  
NIKKTLISLPKSLCDLYHLELLYRPSRLIPDKLNLCKLRLHQLMPSIDGLELSRIPD  
IGRUTLLQRIDSFHLVKQKQKHELRLQLRNMEIGGYLSLRNLNVIGKDEALESKLYQK

Query Match 5.3%; Score 59.4; DB 8; Length 127901;  
Best Local Similarity 73.6%; Pred. No. 1.3e-06;  
Matches 89; Conservative 0; Mismatches 31; Indels 1; Gaps 1;  
Qy 346 GTTTAGAGCAAGTATAAGTCTCAGTCAGCTGGCTATAGAGTGTCCACATCAGCAAA 405  
Db 64764 GCTAAGAGCAAGTATAAGAGCTAATTAGTTGGCGATAGAAGCTCCACATCAGCAT 64823  
Qy 406 TCCTTAAACTGGAGGAGAGAAAGTGGAGTGAAGGGCGTCGGCGCTTCGTCATCG 465  
Db 64824 CTCCTACA-TGGAGGAGAGAGAAATAGGAGAGAGAGTGAAGTGGCGCTTCGAGAGCG 64882  
Qy 466 C 466  
Db 64883 C 64883

RESULT 15

AC137618

LOCUS AC137618 135231 bp DNA linear PLN 03-JUN-2004

## DEFINITION

Oryza sativa (japonica cultivar-group) chromosome 5 clone  
OSUNBa0090H02, complete sequence.

## ACCESSION

AC137618

## VERSION

AC137618.2 GI:48055771

## KEYWORDS

HTG.

## SOURCE

Oryza sativa (japonica cultivar-group)

## ORGANISM

Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.

## REFERENCE

1 (bases 1 to 135231)

## AUTHORS

Chow, T.-Y., Hsing, Y.-I. C., Chen, C.-S., Chen, H.-H., Liu, S.-M.,  
Chao, Y.-T., Chang, S.-J., Chen, H.-C., Chen, S.-K., Chen, F.-R.,  
Chen, Y.-L., Cheng, C.-H., Chung, C.-I., Han, S.-Y., Hsiao, S.-H.,  
Hsiung, J.-N., Hsu, C.-H., Huang, J.-J., Kau, P.-I., Lee, M.-C.,  
Liu, H.-L., Li, Y.-F., Lin, S.-J., Lin, Y.-C., Wu, S.-W., Yu, C.-Y.,  
Yu, S.-W., Wu, H.-P. and Shaw, J.-P.

Oryza sativa BAC OSUNBa0090H02 genomic sequence

## TITLE

Unpublished

## JOURNAL

2 (bases 1 to 135231)

## REFERENCE

Chow, T.-Y. and Hsing, Y.-I. C.

## AUTHORS

Direct Submission

## TITLE

Submitted (27-NOV-2002) Institute of Botany, Academia Sinica, 128,

## JOURNAL

Section 2, Academia Road, Nankang, Taipei 11529, Taiwan

## REFERENCE

3 (bases 1 to 135231)

## AUTHORS

Chow, T.-Y.

## TITLE

Direct Submission

## JOURNAL

Submitted (03-JUN-2004) Institute of Botany, Academia Sinica, 128,

## COMMENT

Section 2, Academia Road, Nankang, Taipei 11529, Taiwan

## FEATURES

On Jun 3, 2004 this sequence version replaced gi:25696256.

## source

Location/Qualifiers

## ORIGIN

1..135231

/organism="Oryza sativa (japonica cultivar-group)"

/mol\_type="genomic DNA"

/cultivar="Nipponbare"

/db\_xref="taxon:39947"

/chromosome="5"

/clone="OSUNBa0090H02"

## Query Match

5.3%; Score 59.4; DB 8; Length 135231;

## Best Local Similarity

73.6%; Pred. No. 1.4e-06;

## Matches

89; Conservative 0; Mismatches 31; Indels 1; Gaps 1;

## Qy

346 GTTTAGAGCAAGTATAAGTCTCAGTCAGCTGGCTATAAGATGTTCCACATCAGCAAA 405

## Db

18217 GCTAAGAGCAAGTATAAGAGCTAATTAGTTGGCGATAAGAAGCTCCACATCAGCAT 18276

## Qy

406 TCCTTAAACTGGAGGAGAGAAAGTAGGAGTGAGAGGGCGTCGGCGCTTCGTCATCG 465

## Db

18277 CTCCTACA-TGGAGGAGAGAGAAATAGGAGAGAGAGTGAAGTGGCGCTTCGAGAGCG 18335

## Qy

466 C 466

## Db

18336 C 18336

Search completed: August 30, 2005, 15:16:34

Job time : 5152 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 30, 2005, 05:58:55 ; Search time 700 Seconds  
(without alignments)  
9395.475 Million cell updates/sec

Title: US-09-979-549-2

Perfect score: 1111

Sequence: 1 gaagtcagaagccgttcag.....ccgagaagagagagagag 1111

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 segs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq\_16Dec04:\*

1: Geneseqn1980s:\*

2: Geneseqn1990s:\*

3: Geneseqn2000s:\*

4: Geneseqn2001as:\*

5: Geneseqn2001bs:\*

6: Geneseqn2002as:\*

7: Geneseqn2002bs:\*

8: Geneseqn2003as:\*

9: Geneseqn2003bs:\*

10: Geneseqn2003cs:\*

11: Geneseqn2003ds:\*

12: Geneseqn2004as:\*

13: Geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1111	100.0	2687	5 AAC84132	Aac84132 Wheat Ta
C 2	57.8	5.2	2000	8 ADA72361	Ada72361 Rice gene
3	47.8	4.3	9538	6 AAS20135	Aas20135 Genomic D
4	47.8	4.3	9539	6 AAS20019	Aas20019 Genomic D
C 5	45.8	4.1	1976	12 ADJ44967	Adj44967 Plant cDN
C 6	41	3.7	1999	10 ADC08428	Adc08428 Rice DNA
C 7	39.8	3.6	2000	8 ADA71571	Ada71571 Rice gene
C 8	39.4	3.5	69300	6 AAD38804	Aad38804 BAC clone
C 9	38	3.4	24333	12 ADJ12497	Adj12497 DNA fragm
C 10	38	3.4	24333	12 ADJ12606	Adj12606 DNA fragm
11	37.2	3.3	88191	8 ABX14763	Abx14763 Genomic D
12	36.6	3.3	2000	8 ADA71938	Ada71938 Rice gene
13	36.2	3.3	1898	11 ACN91201	Acn91201 Breast ca
C 14	36.2	3.3	2000	10 ADC08414	Adc08414 Rice DNA
C 15	36.2	3.3	5955	12 ADH22268	Adh22268 Rice PONG
C 16	35.8	3.2	460	3 AAC28189	Aac28189 Human sec
C 17	35.2	3.2	2000	8 ADA71938	Ada71938 Rice gene
C 18	35	3.2	704	6 ABQ32010	Abq32010 Oligonuc
19	35	3.2	704	6 ABQ32011	Abq32011 Oligonuc
20	34.8	3.1	7525	4 AAS01192	Aas01192 Fertiliza

21	34.8	3.1	7525	10 ADC29908	Adc29908 Fertiliza
22	34.8	3.1	13031	11 ADM39442	Adm39442 Maize ZnF
C 23	34.6	3.1	837	6 ABQ35188	Abq35188 Oligonuc
24	34.6	3.1	837	6 ABQ35189	Abq35189 Oligonuc
C 25	34.2	3.1	8649	4 AAS46329	Aas46329 Tumour eu
C 26	34.2	3.1	8649	6 ABK31246	Abk31246 Signal tr
C 27	34.2	3.1	8649	6 ABL70201	Ab170201 Chemical
C 28	34.2	3.1	8649	6 ABL70201	Ab170201 Chemical
29	34.2	3.1	118931	11 ACN44586	Acn44586 Human gen
30	34	3.1	561	10 ADK52808	Adk52808 Plant DNA
31	34	3.1	661	13 ADR64189	Adr64189 Cotton cD
C 32	33.8	3.0	3428	6 ABI99351	Abi99351 Mouse isc
C 33	33.8	3.0	6523	10 ADE84216	Ad84216 Human Lym
C 34	33.8	3.0	6523	13 ADS89620	Ad89620 Oligonuc
C 35	33.8	3.0	59458	13 ABD33282	Abd33282 Murine ca
C 36	33.8	3.0	59458	13 ADR67037	Adr67037 Mouse can
37	33.6	3.0	484	5 ABV59185	Abv59185 Human pro
38	33.4	3.0	423	13 ADR60416	Adr60416 Cotton cD
39	33.4	3.0	527	6 ABQ92787	Abq92787 Triticum
C 40	33.4	3.0	2467	6 ABA90875	Ab90875 Bacillus
C 41	33.4	3.0	5642	4 AAK69744	Aak69744 Human imm
C 42	33.4	3.0	5642	5 AAS34694	Aas34694 Human DNA
C 43	33.2	3.0	358	5 ABV15204	Abv15204 Human pro
44	33.2	3.0	513	6 ABQ33211	Abq33211 Oligonuc
C 45	33.2	3.0	513	6 ABQ33210	Abq33210 Oligonuc

ALIGNMENTS

RESULT 1

AAC84132

ID AAC84132 standard; DNA; 2687 BP.

XX AAC84132;

AC AAC84132;

DT 09-APR-2001 (first entry)

DE Wheat TaTrxh2 gene.

XX Wheat; TaTrxh2; thioredoxin; promoter; gene expression; transgenic plant;

KW monocotyledon; plant cell; seed; amylaceous albumen; db.

XX Triticum aestivum.

OS Triticum aestivum.

XX Key Location/Qualifiers

FT promoter 1..1111

FT /tag= a

FT /note= "promoter region is specifically claimed"

FT misc\_signal 550..558

FT /tag= b

FT /label= Gibberellic\_acid\_response\_element

FT /function= "regulates gene expression in response to gibberellic acid"

FT misc\_signal 561..569

FT /tag= c

FT /label= Gibberellic\_acid\_response\_element

FT /function= "regulates gene expression in response to gibberellic acid"

FT protein\_bind 599..708

FT /tag= d

FT /bound\_moiety= "GCNA-like protein"

FT /label= GCNA-like\_box

FT protein\_bind 860

FT /tag= e

FT /bound\_moiety= "leucine zipper proteins"

FT /label= bzip\_motif

FT misc\_signal 867..883

FT /tag= f

FT /function= "possible regulator of gene expression in response to abscisic acid"

FT /note= "present in thioredoxin h gene promoter sequences from tobacco and rice"



QY 961 TAGCGAGCGCATAAATTCTGATTCTCTGCTGCTGCGGCAATTTATCTTTGGGAGGC 1020  
|||||  
Db 961 TAGCGAGCGCATAAATTCTGATTCTCTGCTGCTGCGGCAATTTATCTTTGGGAGGC 1020  
|||||  
QY 1021 GGGCGGGATTGGAGAGAGAGCCACAGGCGACACAAAGTGGCGGTGAGAAATCAACA 1080  
|||||  
Db 1021 GGGCGGGATTGGAGAGAGAGCCACAGGCGACACAAAGTGGCGGTGAGAAATCAACA 1080  
|||||  
QY 1081 AGCGGTGCTTCCGAGAGAGAGAGAGAGAG 1111  
|||||  
Db 1081 AGCGGTGCTTCCGAGAGAGAGAGAGAGAG 1111  
|||||  
RESULT 2  
ADA72361/c  
ID ADA72361 standard; DNA; 2000 BP.  
XX  
AC ADA72361;  
XX  
AC ADA72361;  
XX  
DT 20-NOV-2003 (first entry)  
XX  
XX Rice gene, SEQ ID 5686.  
DE  
XX Plant; bacterial infection; fungal infection; viral infection; rice;  
KW gene; ds.  
XX  
XX Oryza sativa.  
OS  
XX WO2003000898-A1.  
PN  
XX  
XX 03-JAN-2003.  
PD  
XX  
XX 22-JUN-2001; 2001WO-IB001105.  
PF  
XX  
XX 22-JUN-2001; 2001WO-IB001105.  
PR  
XX  
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
PA  
XX  
XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
XX  
XX WPI; 2003-175290/17.  
DR  
XX  
XX Identifying at least one gene involved in plant resistance or response to  
PT pathogenic infection for conferring resistance or tolerance to a plant to  
PT bacterial, fungal or viral infection by determining or detecting plant  
PT gene expression.  
XX  
XX Claim 27; SEQ ID NO 5686; 899pp; English.  
PS  
XX  
XX The present invention relates to a method (M1) for identifying genes  
CC involved in plant resistance or response to pathogenic infection. M1  
CC comprises identifying a gene whose expression is significantly altered in  
CC the incompatible interaction of plant gene expression relative to  
CC expression of the gene in an uninfected plant, in a mutant plant that  
CC does not express a gene associated with response to pathogenic infection,  
CC or in a corresponding incompatible or compatible interaction. (M1) is  
CC useful for conferring resistance to resistance or tolerance to a plant to  
CC bacterial, fungal or viral infection. The present sequence was used to  
CC illustrate the invention.  
XX  
XX Sequence 2000 BP; 557 A; 446 C; 388 G; 609 T; 0 U; 0 Other;  
SQ  
Query Match 5.2%; Score 57.8; DB 8; Length 2000;  
Best Local Similarity 64.0%; Pred. No. 1.4e-07;  
Matches 103; Conservative 0; Mismatches 57; Indels 1; Gaps 1;  
QY 329 GCGGCGAGTGGCTTCCCGTTTAGAGCAAGTATTAAGTCTAGTCAGCTGGCTATAAGA 388  
|||||  
Db 404 GCGGCTAGGAGCAGTGTGTAGAGAGAGTACATAAG-CTTAGTCAGCTGGCGGCAAA 346  
|||||  
QY 389 TGTTCACATCAGCAATCCTTAACTGGAGGAGAAAGATGAGGTAGGAGGCGGT 448  
|||||

Db 345 TCATCCACGTAATAAATCTTCCCGTGAAGAGAGAGAGGGGGGAAGAGAGCAACC 286  
|||||  
QY 449 CGGCGCTTCGTCGAATCGCTAGCATAGACACAAGCTCCCATG 489  
|||||  
Db 285 GGACGCTTATTAAAGCGCTAGCGCTAACACATAAACAATG 245  
|||||  
RESULT 3  
AAS20135  
ID AAS20135 standard; cDNA; 9538 BP.  
XX  
AC AAS20135;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
XX Genomic DNA encoding fructosyl transferase homologue, Lp6SFT1 version 2.  
DE  
XX Fructosyl transferase homologue; Lp6SFT1; perennial ryegrass; fescue;  
KW fructan biosynthesis; forage; turf grass improvement;  
KW dry matter digestibility; herbage quality; palatability; cold tolerance;  
KW drought tolerance; tiller survival; plant persistence; abiotic stress;  
KW low calorie sweetener; ds.  
XX  
XX Lolium perenne.  
OS  
XX  
XX Location/Qualifiers  
FH Key  
CDS 5702..9457  
FT /\*tag= a "Lp6SFT1"  
FT /product= "Lp6SFT1"  
FT /note= "Lolium perenne fructosyl transferase"  
FT /partial  
FT /note= "No stop codon given"  
FT /transl\_except= (pos:7176..7186, aa:WRTV)  
FT exon 5702..5112  
FT /\*tag= b  
FT /number= 1  
FT intron 6113..6327  
FT /\*tag= c  
FT /number= 1  
FT exon 6328..6336  
FT /\*tag= d  
FT /number= 2  
FT intron 6337..6773  
FT /\*tag= e  
FT /number= 2  
FT exon 6774..7612  
FT /\*tag= f  
FT /number= 3  
FT intron 7613..9055  
FT /\*tag= g  
FT /number= 3  
FT exon 9056..9457  
FT /\*tag= h  
FT /number= 4  
XX WO200195691-A2.  
PD  
XX 20-DEC-2001.  
XX  
XX 14-JUN-2001; 2001WO-AU000705.  
PF  
XX  
XX 14-JUN-2000; 2000AU-00008155.  
XX  
XX (VICT-) STATE VICTORIA DEPT NATURAL RES & ENVIRO.  
PA (UYAD-) UNIV ADELAIDE.  
PA (ITWA-) INT MAIZE & WHEAT IMPROVEMENT CENT.  
PA (SAUS-) STATE SOUTH AUSTRALIA SOUTH AUSTRALIAN R.  
XX (UYSC-) UNIV SOUTHERN CROSS.  
PI Spangenberg GC, Lidgett AJ, Johnson XA, Terdich K;  
XX WPI; 2002-106436/14.  
DR





Db 7990 AGAGCAAGTACAATAAGATCTAGCTCAGCTCGCTACAGGATTAATAATATTTGTGT 8049  
Qy 410 TAAACTGGAGGAGAGAAAGTAGGAGTGAGAGGGCGTCGGCGCTTCGGTCAATCGCTAG 469  
Db 8050 CTAGTTGGAGGAGATAGGAGGAGAGAGA-ATGTGATGATGCTCTTATCCAGAGCTAG 8108  
Qy 470 CGATAGCACAAAGTCCCATCGAA 492  
Db 8109 CTCTAGCACGCTGCTCTTAGGCAA 8131

RESULT 5  
ADJ44967/c  
ID ADJ44967 standard; cdna; 1976 BP.  
XX AC ADJ44967;  
XX 06-MAY-2004 (first entry)  
XX DE Plant cdna #5967.  
XX Plant; gene; ss; transcription; plant genome augmentation; cereal;  
KW soybean; alfalfa; sunflower; canola; cotton; peanut; tobacco; sugar beet;  
KW maize; barley; sorghum; rice; wheat; crop plant; insecticide resistance;  
KW stress tolerance; salt tolerance; cold tolerance; drought tolerance;  
KW plant nutrition; apical dominance; dwarfism; early flowering; antiviral;  
KW antifungal.  
XX Eukaryota.  
XX US2004016025-A1.  
XX 22-JAN-2004.  
XX 26-SEP-2002; 2002US-00260238.  
XX 26-SEP-2001; 2001US-0325277P.  
XX 26-SEP-2001; 2001US-0325448P.  
XX 04-APR-2002; 2002US-0370620P.  
XX (BUDW/) BUDWORTH P.  
XX (MOUG/) MOUGHAMER T.  
XX (BRIG/) BRIGGS S P.  
XX (COOP/) COOPER B.  
XX (GLAZ/) GLAZEBROOK J.  
XX (GOFF/) GOFF S A.  
XX (KATA/) KATAGIRI F.  
XX (KREP/) KREPS J.  
XX (PROV/) PROVART N.  
XX (RICK/) RICHE D.  
XX (ZHUT/) ZHU T.  
XX Budworth P, Moughamer T, Briggs SP, Cooper B, Glazebrook J;  
PI Goff SA, Katagiri F, Kreps J, Provart N, Ricke D, Zhu T;  
XX WPI; 2004-190374/18.  
XX New rice promoter, useful for manipulating crop plants to alter or  
PT improve phenotypic characteristics, e.g. produce large quantities of oil  
PT or proteins, resistance to insecticides, virus or fungi, stress tolerance  
PT or high nutritional value.  
XX Claim 26; SEQ ID NO 5967; 230pp; English.

XX The invention relates to plant nucleotide sequences that direct seed-,  
CC leaf- and/or stem-, panicle-, root- or pollen-specific or -preferential  
CC or constitutive transcription of an operatively linked nucleic acid  
CC segment. The invention also relates to a method for augmenting a plant  
CC genome and a method of identifying a gene, where its expression is  
CC altered in the seed, leaf, stem, panicle, pollen, root or is constitutive  
CC in a plant cell. The plant is a cereal, e.g. soybean, alfalfa, sunflower,  
CC canola, cotton, peanut, tobacco or sugar beet, preferably maize, barley,

CC sorghum, rice or wheat. The polynucleotides and the polypeptides they  
CC encode are useful for manipulating crop plants to alter or improve  
CC phenotypic characteristics, to produce large quantities of oil or  
CC proteins, to incur resistance to insecticides, viruses or fungi, and to  
CC incur stress tolerance (e.g. salt, cold or drought) to ensure the plants  
CC have a high nutritional value with reduced apical dominance or dwarfism,  
CC early flowering or altered metabolic pathways. This sequence represents a  
CC plant nucleic acid of the invention. Note: The sequence data for this  
CC patent did not form part of the printed specification but was obtained in  
CC electronic format directly from USPTO at seqdata.uspto.gov/sequence.html.  
XX  
SQ Sequence 1976 BP; 489 A; 464 C; 447 G; 576 T; 0 U; 0 Other;

Query Match 4.1%; Score 45.8; DB 12; Length 1976;  
Best Local Similarity 58.8%; Pred. No. 0.0014;  
Matches 97; Conservative 0; Mismatches 67; Indels 1; Gaps 1;  
Qy 348 TTAGAGCAAGTATATAAGTCTCTAGTCAGCTGGCTATAGTGTTCACATCGCAATC 407  
Db 696 TAAGAGCAAGTATAGTAAATGTACGTGCGGAAATCAACCGCTCGTCTCATCCAAAT 637  
Qy 408 CTTAAACTGGAGGAGAGAAAGTAGGAGTGAGAGGGCGTTCGGCTTCGTCGAATCGCT 467  
Db 636 C-CAAGCTGGAGGAGGAGAAACCAAGAAAGAGAGAAATGGCGAGTGGCGAGTCGCC 578  
Qy 468 AGCGATAGCACAAAGCTCCCATCGAATCGAGCCCAACATGCAACCCG 512  
Db 577 GGCTCGAAGCGTCTAGTACCGGAGGCAAAACGCCGCTCCCAACTCG 533

RESULT 6  
ADC08428/c  
ID ADC08428 standard; DNA; 1999 BP.  
XX AC ADC08428;  
XX 18-DEC-2003 (first entry)  
XX DE Rice DNA sequence Seq ID733 related to grain filling.  
XX plant biotechnology; carbohydrate synthesis; carbohydrate metabolism;  
KW carbohydrate degradation; carbohydrate; plant grain; grain filling; corn;  
KW tomato; banana; canola; cotton; peanut; sorghum; tobacco; sugarbeet;  
KW wheat; rice; protein; oil; starch; fibre; moisture content; cereal grain;  
KW gene; ds; plant.  
XX Oryza sativa.  
XX WO2003000905-A2.  
XX 03-JAN-2003.  
XX 21-JUN-2002; 2002WO-IB002450.  
XX 22-JUN-2001; 2001US-0300112P.  
XX 26-SEP-2001; 2001US-0325277P.  
XX 20-DEC-2001; 2001US-0342327P.  
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX Zhu T, Cheng W, Briggs S, Cooper B, Goff SA, Moughamer T;  
PI Glazebrook J, Katagiri F, Kreps J, Provart N, Ricke D;  
XX WPI; 2003-229341/22.  
XX New plant genes encoding polypeptides having an activity involved in or  
PT associated with the synthesis, metabolism or degradation of carbohydrates  
PT in the plant grain useful in generating plants having improved  
PT nutritional properties.  
XX Disclosure; SEQ ID NO 733; 130pp; English.  
XX This invention, in the area of plant biotechnology, relates to novel

CC polynucleotides comprising a nucleotide sequence encoding a protein which  
CC is involved in or associated with the synthesis, metabolism or  
CC degradation of carbohydrates in the plant grain and the expression of  
CC which is up-regulated during grain filling. The plant is selected from  
CC corn, tomato, banana, canola, cotton, peanut, sorghum, tobacco,  
CC sugarbeet, wheat, and rice. The invention may be useful for the  
CC improvement of protein, oil, starch, fibre and moisture content of the  
CC cereal grains. In addition, carbohydrate levels may be modified to a more  
CC desirable level using the present invention. The present sequence is a  
CC DNA sequence of a rice gene promoter. Note: the sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/publishedpct\_sequences.

XX  
SQ Sequence 1999 BP; 624 A; 366 C; 428 G; 580 T; 0 U; 1 Other;  
Query Match 3.7%; Score 41; DB 10; Length 1999;  
Best Local Similarity 72.6%; Pred. No. 0.054;  
Matches 53; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 628 TATAGCTGACCTGGCAGTGTATAGAGCGCGGCTCTTCTATTAGCTTTGCTCTT 687  
Db 282 TATAGATGAGATGTGTAGTAGTATTCAGCCAGCGCGCTAAAGTATTAGCCTTGCTCTT 223

QY 688 ATGGCTACATCTG 700  
Db 222 AGGAGGGGATCTG 210

RESULT 7  
ADA71571/c  
ID ADA71571 standard; DNA; 2000 BP.  
XX  
AC ADA71571;  
XX  
XX 20-NOV-2003 (first entry)  
XX  
XX Rice gene, SEQ ID 4894.  
XX  
XX Plant; bacterial infection; fungal infection; viral infection; rice;  
XX gene; ds.  
XX  
XX Oryza sativa.  
XX  
XX WO2003000898-A1.  
XX  
XX 03-JAN-2003.  
XX  
XX 22-JUN-2001; 2001WO-IB001105.  
XX  
XX 22-JUN-2001; 2001WO-IB001105.  
XX  
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX  
XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
XX Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
XX WPI; 2003-175290/17.  
XX  
XX Identifying at least one gene involved in plant resistance or response to  
XX pathogenic infection for conferring resistance or tolerance to a plant to  
XX bacterial, fungal or viral infection by determining or detecting plant  
XX gene expression.  
XX  
XX Claim 27; SEQ ID NO 4894; 899pp; English.  
XX  
XX The present invention relates to a method (M1) for identifying genes  
XX involved in plant resistance or response to pathogenic infection. M1  
XX comprises identifying a gene whose expression is significantly altered in  
XX the incompatible interaction of plant gene expression relative to  
XX expression of the gene in an uninfected plant, in a mutant plant that  
XX does not express a gene associated with response to pathogenic infection,  
XX or in a corresponding incompatible or compatible interaction. (M1) is

CC useful for conferring resistance to resistance or tolerance to a plant to  
CC bacterial, fungal or viral infection. The present sequence was used to  
CC illustrate the invention.

XX  
SQ Sequence 2000 BP; 585 A; 533 C; 403 G; 479 T; 0 U; 0 Other;  
Query Match 3.6%; Score 39.8; DB 8; Length 2000;  
Best Local Similarity 67.5%; Pred. No. 0.13;  
Matches 56; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 329 GCGGCGAGTGGCTTCCCGTTTACAGCAAGTATATAAGTCTTAGTCAGCTGGCTATAAGA 368  
Db 91 GCAGCAAGCTTTGTGTCTCTAAGAGCAAGTATAGTAAGTCTTAGTCAGCGGGAACCG 32

QY 389 TGTTCACATCAGCAATCCTTA 411  
Db 31 TTTCCACGTCATCAAAATCCTA 9

RESULT 8  
AAD38804  
ID AAD38804 standard; DNA; 69300 BP.  
XX  
AC AAD38804;  
XX  
XX 23-SEP-2002 (first entry)  
XX  
XX BAC clone E2P5 from rice variety CO39.  
XX  
XX Rice; chromosome 11; Indica rice cultivar; CO39; avirulence gene;  
XX AVR1-CO39; rice blast pathogen; transgenic plant; plant breeding;  
XX resistance; agricultural; horticultural; plant protectant; ds.  
XX  
XX Oryza sativa.  
XX  
XX WO200234927-A2.  
XX  
XX 02-MAY-2002.  
XX  
XX 19-OCT-2001; 2001WO-US046331.  
XX  
XX 20-OCT-2000; 2000US-0242313P.  
XX  
XX 09-JUL-2001; 2001US-0303897P.  
XX  
XX (WISC ) WISCONSIN ALUMNI RES FOUND.  
XX  
XX (USDA ) US DEPT OF AGRICULTURE.  
XX  
XX Leong SA, Farman ML, Chauhan RS, Durfee TJ;  
XX WPI; 2002-471442/50.  
XX  
XX New PiCO39(t) polynucleotides from chromosome 11 of Indica rice cultivar  
XX CO39 useful for conferring or improving resistance of plants to strains  
XX of Magnaporthe grisea or other plant pathogens with avirulence gene AVR1-  
XX CO39.  
XX  
XX Example 4; Page 110-142; 175pp; English.  
XX  
XX The invention relates to a polynucleotide isolated from chromosome 11 of  
XX Indica rice cultivar CO39, flanked by marker R2316 and RG1094 comprising  
XX one or more genes that confer resistance to strains of Magnaporthe grisea  
XX having avirulence gene AVR1-CO39. The Pi-CO39(t) polynucleotides are  
XX useful for conferring or improving resistance of plants to strains of the  
XX rice blast pathogen, Magnaporthe grisea and other pathogens. The Pi-  
XX CO39(t) nucleic acids may be used as probes to detect the presence of  
XX and/or expression of Pi-CO39(t) genes, and to produce large quantities of  
XX pure Pi-CO39(t) proteins. Purified gene products of Pi-CO39(t) are useful  
XX in producing polyclonal or monoclonal antibodies useful as sensitive  
XX detection reagents for the presence and accumulation of Pi-CO39(t)  
XX polypeptides. The transgenic plants are useful for plant breeding or  
XX directly in agricultural or horticultural applications. The present  
XX sequence is BAC clone E2P5 from rice variety CO39

SQ Sequence 69300 BP; 19720 A; 15094 C; 15006 G; 19380 T; 0 U; 100 Other;  
Query Match 3.5%; Score 39.4; DB 6; Length 69300;  
Best Local Similarity 65.2%; Pred. No. 1.5;  
Matches 58; Conservative 0; Mismatches 31; Indels 0; Gaps 0;  
QY 601 CAGCCAGTTTATATATAACAGGCTATATAGCTGACCTGGAGTGTATAGAGCCGGCA 660  
DB 35667 CAGCCGCGCTTATATATAGTGAAGGCTTTAGGTGATGTGTCTACTAGTATAACGCCCACT 35726  
QY 661 GCCGGCTCTCTTATATAGCTTTGCTCTTAT 689  
DB 35727 GCTGCGCTTCTTATATCTTGTCTTAT 35755  
RESULT 9  
ADJ12497/c  
ID ADJ12497 standard; DNA; 24333 BP.  
AC ADJ12497;  
XX  
DT 20-MAY-2004 (first entry)  
XX  
DE DNA fragment of a BAC clone that encodes a human secreted protein Seq351.  
XX  
KW human; secreted; cancer; haematopoietic disease; anaemia;  
KW multiple myeloma; reproductive system disorder; prostatitis;  
KW inguinal hernia; musculoskeletal disease; systemic lupus erythematosus;  
KW gout; cardiovascular disease; arrhythmia; hypernatraemia; fetal disease;  
KW fetal alcohol syndrome; Down's syndrome; excretory disease;  
KW urinary incontinence; renal disorder; neural; sensory disease;  
KW Alzheimer's disease; meningitis; endocrine disease; diabetes;  
KW occupational lung disease; endocrine disease; diabetes;  
KW glomerulonephritis; digestive disease; portal hypertension;  
KW irritable bowel syndrome; epithelial disease; scleroderma;  
KW epidermolysis bullosa; cytostatic; antianemic; antiarthritic;  
KW antidiabetic; anti-HIV; immunosuppressive; antinflammatory;  
KW antipsoriatic; antibacterial; osteopathic; dermatological; antigitout;  
KW immunomodulator; antiarrhythmic; cardiant; neutropic; antilipemic;  
KW nephrotropic; uropathic; neuroprotective; antiparkinsonian; tranquilizer;  
KW antidiabetic; anabolic; hypertensive; vulnerary; ds.  
XX  
OS Homo sapiens.  
XX  
PN US2004010132-A1.  
XX  
PD 15-JAN-2004.  
XX  
PF 30-OCT-2001; 2001US-00984429.  
XX  
PR 09-OCT-1997; 97US-0061463P.  
PR 09-OCT-1997; 97US-0061527P.  
PR 09-OCT-1997; 97US-0061529P.  
PR 09-OCT-1997; 97US-0061532P.  
PR 09-OCT-1997; 97US-0061536P.  
PR 09-OCT-1997; 97US-0071498P.  
PR 08-OCT-1998; 98WO-US021142.  
PR 08-APR-1999; 99US-00288143.  
PR 01-NOV-2000; 2000US-0244591P.  
XX  
PA (ROSE/) ROSEN C A.  
PA (BREW/) BREWER L A.  
PA (DUAN/) DUAN R D.  
PA (RUBE/) RUBEN S M.  
PA (FLOR/) FLORENCE K A.  
PA (GREE/) GREENE J M.  
PA (YOUN/) YOUNG P E.  
PA (FERR/) FERRIE A M.  
PA (YUGG/) YU G.  
PA (FLOR/) FLORENCE C.  
PA (EBNE/) EBNER R.  
PA (OLSE/) OLSEN H.  
XX

PI Rosen CA, Brewer LA, Duan RD, Ruben SM, Florence KA, Greene JM;  
PI Young PE, Ferrie AM, Yu G, Florence C, Ebner R, Olsen H;  
XX WPI; 2004-090518/09.  
DR  
XX  
XX New isolated nucleic acids and polypeptides, useful for diagnosing,  
PT treating, preventing or ameliorating diseases or disorders e.g. cancer,  
PT anaemia, arthritis, asthma, inflammatory bowel disease or Alzheimer's  
PT disease.  
XX  
PS Disclosure; SEQ ID NO 351; 286pp; English.  
XX  
XX This invention relates to novel polynucleotides encoding human secreted  
CC proteins. Specifically, it refers to the vectors, host cells, recombinant  
CC and synthetic methods for producing human polynucleotides, polypeptides  
CC and antibodies. Furthermore, it relates to screening methods to identify  
CC agonists and antagonists that can be used to inhibit or enhance the  
CC production and function of the secreted proteins. The present invention  
CC describes these compositions as useful for diagnosing, treating or  
CC preventing disorders such as cancer, haematopoietic diseases including  
CC anaemia and multiple myeloma, reproductive system disorders including  
CC prostatitis and inguinal hernia, musculoskeletal diseases including  
CC systemic lupus erythematosus and gout, cardiovascular disease including  
CC arrhythmia and hypernatraemia, mixed fetal diseases including fetal  
CC alcohol syndrome and Down's syndrome, excretory diseases including  
CC urinary incontinence and renal disorders, neural or sensory disease  
CC including Alzheimer's disease and meningitis, respiratory disease  
CC including emphysema and occupational lung disease, endocrine diseases  
CC including diabetes and glomerulonephritis, digestive diseases including  
CC portal hypertension and irritable bowel syndrome and connective tissue or  
CC epithelial diseases including scleroderma and epidermolysis bullosa. As  
CC such, there are various activities such as cytostatic, antianemic,  
CC antiarthritic, antiasthmatic, anti-HIV, immunosuppressive,  
CC antinflammatory, antipsoriatic, antibacterial, osteopathic,  
CC dermatological, antigitout, immunomodulator, antiarrhythmic, cardiant,  
CC neutropic, antilipemic, nephrotropic, uropathic, neuroprotective,  
CC antiparkinsonian, tranquilizer, antidiabetic, anabolic, hypertensive and  
CC vulnerary. This polynucleotide is a DNA fragment of a BAC clone that  
CC encodes a human secreted protein of the invention. NOTE: This sequence  
CC does not appear in the printed specification but has been obtained in  
CC electronic format from the US patent office at the following web site  
CC www.seqdata.uspto.gov/sequence.html; Document ID: 20040010132.  
XX  
SQ Sequence 24333 BP; 6183 A; 6524 C; 6176 G; 5450 T; 0 U; 0 Other;  
Query Match 3.4%; Score 38; DB 12; Length 24333;  
Best Local Similarity 55.2%; Pred. No. 2.3;  
Matches 74; Conservative 0; Mismatches 60; Indels 0; Gaps 0;  
QY 179 ATTTTATGGCGATGGCGTTCAGCGCTTTATCTAGGCGTCTGGGAGGTACATTGAAGATG 238  
DB 8903 ATCTGTTGGGCTGGTGACAGGCGCTGTCTGCAGAGGACTTGTCTAAGAAGATGAAGATG 8844  
QY 239 TGCCACCACTCCAAACCGACACCTGTATCTGAGCATGCTCTCTCTCTTCATG 298  
DB 8843 CAACACCCAACTCTCCCGCTGGCTAGATGGGAGCATGGAGCAACCAACCTTTGCTG 8784  
QY 299 CTCTCCCTTTGGGTG 312  
DB 8783 CTGCTGTGGGGTG 8770  
RESULT 10  
ADJ12606/c  
ID ADJ12606 standard; DNA; 24333 BP.  
XX  
AC ADJ12606;  
XX  
DT 20-MAY-2004 (first entry)  
XX  
DE DNA fragment of a BAC clone that encodes a human secreted protein Seq460.  
XX human; secreted; cancer; haematopoietic disease; anaemia;  
XX

KW multiple myeloma; reproductive system disorder; prostatitis;  
 KW inguinal hernia; musculoskeletal disease; systemic lupus erythematosus;  
 KW gout; cardiovascular disease; arrhythmia; hypernatremia; fetal disease;  
 KW fetal alcohol syndrome; Down's syndrome; excretory disease;  
 KW urinary incontinence; renal disorder; neural; sensory disease;  
 KW Alzheimer's disease; meningitis; respiratory disease; emphysema;  
 KW occupational lung disease; endocrine disease; diabetes;  
 KW glomerulonephritis; digestive disease; portal hypertension;  
 KW irritable bowel syndrome; epithelial disease; scleroderma;  
 KW epidermolysis bullosa; cytostatic; antianemic; antiarthritic;  
 KW antiasthmatic; anti-HIV; immunosuppressive; antiinflammatory;  
 KW antiparasitic; antibacterial; osteopathic; dermatological; angiot; angiot;  
 KW immunomodulator; antiarrhythmic; cardiant; neurotropic; antilipemic;  
 KW nephrotropic; uropathic; neuroprotective; antiparkinsonian; tranquilizer;  
 KW antidiabetic; anabolic; hypertensive; vulnerary; ds.  
 OS Homo sapiens.  
 XX US2004010132-A1.  
 XX 15-JAN-2004.  
 XX 30-OCT-2001; 2001US-00984429.  
 XX 09-OCT-1997; 97US-0061463P.  
 XX 09-OCT-1997; 97US-0061527P.  
 XX 09-OCT-1997; 97US-0061529P.  
 XX 09-OCT-1997; 97US-0061532P.  
 XX 09-OCT-1997; 97US-0061536P.  
 XX 09-OCT-1997; 97US-0071498P.  
 XX 08-OCT-1998; 98WO-US021142.  
 XX 08-APR-1999; 99US-00288143.  
 XX 01-NOV-2000; 2000US-0244591P.  
 XX (ROSE/) ROSEN C A.  
 XX (BREW/) BREWER L A.  
 XX (DUAN/) DUAN R D.  
 XX (RUBE/) RUBEN S M.  
 XX (FLOR/) FLORENCE K A.  
 XX (GREEN/) GREENE J M.  
 XX (YOUN/) YOUNG P E.  
 XX (FERR/) FERRIE A M.  
 XX (YUGG/) YU G.  
 XX (FLOR/) FLORENCE C.  
 XX (BENE/) BENNER R.  
 XX (OLSE/) OLSEN H.  
 XX Rosen CA, Brewer LA, Duan RD, Ruben SM, Florence KA, Greene JM;  
 XX Young PE, Ferrie AW, Yu G, Florence C, Ebner R, Olsen H;  
 XX WPI; 2004-090518/09.  
 XX New isolated nucleic acids and polypeptides, useful for diagnosing,  
 XX treating, preventing or ameliorating diseases or disorders e.g. cancer,  
 XX anemia, arthritis, asthma, inflammatory bowel disease or Alzheimer's  
 XX disease.  
 XX Disclosure; SEQ ID NO 460; 286pp; English.  
 XX This invention relates to novel polynucleotides encoding human secreted  
 XX proteins. Specifically, it refers to the vectors, host cells, recombinant  
 XX and synthetic methods for producing human polynucleotides, polypeptides  
 XX and antibodies. Furthermore, it relates to screening methods to identify  
 XX agonists and antagonists that can be used to inhibit or enhance the  
 XX production and function of the secreted proteins. The present invention  
 XX describes these compositions as useful for diagnosing, treating or  
 XX preventing disorders such as cancer, haematopoietic diseases including  
 XX anaemia and multiple myeloma, reproductive system disorders including  
 XX prostatitis and inguinal hernia, musculoskeletal diseases including  
 XX systemic lupus erythematosus and gout, cardiovascular disease including  
 XX arrhythmia and hypernatremia, mixed fetal diseases including fetal  
 XX alcohol syndrome and Down's syndrome, excretory diseases including  
 XX urinary incontinence and renal disorders, neural or sensory disease

CC including Alzheimer's disease and meningitis, respiratory disease  
 CC including emphysema and occupational lung disease, endocrine diseases  
 CC including diabetes and glomerulonephritis, digestive diseases including  
 CC portal hypertension and irritable bowel syndrome and connective tissue or  
 CC epithelial diseases including scleroderma and epidermolysis bullosa. As  
 CC such, there are various activities such as cytostatic, antianemic,  
 CC antiarthritic, antiasthmatic, anti-HIV, immunosuppressive,  
 CC antiinflammatory, antiparasitic, antibacterial, osteopathic,  
 CC dermatological, angiot, angiot, immunomodulator, antiarrhythmic, cardiant,  
 CC neurotropic, antilipemic, nephrotropic, uropathic, neuroprotective,  
 CC antiparkinsonian, tranquilizer, antidiabetic, anabolic, hypertensive and  
 CC vulnerary. This polynucleotide is a DNA fragment of a BAC clone that  
 CC encodes a human secreted protein of the invention. NOTE: This sequence  
 CC does not appear in the printed specification but has been obtained in  
 CC electronic format from the US patent office at the following web site  
 CC www.seqdata.uspto.gov/sequence.html; Document ID: 20040010132.  
 XX  
 SQ Sequence 24333 BP; 6183 A; 6524 C; 6176 G; 5450 T; 0 U; 0 Other;  
 Query Match 3.4%; Score 38; DB 12; Length 24333;  
 Best Local Similarity 55.2%; Pred. No. 2.3;  
 Matches 74; Conservative 0; Mismatches 60; Indels 0; Gaps 0;  
 QY 179 ATTTTATGGCGATGGCGTCAAGCGTTTATCTAGGCGCTCTGGAGGCTACATTTGAAGATG 238  
 DB 8903 ATCTGGTGGGCTGGTGACAGGCGCTCTCTGACAGAGGACTTGTCTAAGAAGATGAAGATG 8844  
 QY 239 TGGCAGCAACTCCAAACCGACCAACCTGTATCTGAGCATGCGCTCTCTCTTCATG 298  
 DB 8843 CAACACCCAACTCTCCCGCTGGCTAGATGGAGCATGGAGCAACCAACCTTTGCTG 8784  
 QY 299 CTCTCCTTTGGGGTG 312  
 DB 8783 CTCTGCTGGGGTG 8770  
 RESULT 11  
 ABX14763  
 ID ABX14763 standard; DNA; 88191 BP.  
 AC ABX14763;  
 DT 01-APR-2003 (first entry)  
 XX  
 DE  
 XX  
 XX Genomic DNA encoding novel human ras-like protein.  
 XX Ras-like protein; pharmacogenomic analysis; tissue typing; human;  
 KW transgenic; ribozyme design; gene therapy; alternative splice form; gene;  
 KW ds.  
 XX  
 OS Homo sapiens.  
 XX  
 XX US2002132291-A1.  
 XX 19-SEP-2002.  
 XX 07-MAR-2001; 2001US-00799799.  
 XX 07-MAR-2001; 2001US-00799799.  
 XX (YEJ/) YE J.  
 XX (KETC/) KETCHUM K A.  
 XX (DFRA/) DI FRANCESCO V.  
 XX (BEAS/) BEASLEY E M.  
 XX Ye J, Ketchum KA, Di Francesco V, Beasley EM;  
 XX WPI; 2003-174080/17.  
 DR P-PSDB; ABG72926.  
 XX Novel isolated human Ras-like protein useful for treating disorders  
 XX characterized by absence of, inappropriate or unwanted expression of Ras-  
 XX like protein, and as immunogens to raise antibodies.

XX Claim 4; Fig 3; 106pp; English.

XX The invention describes an isolated novel human Ras-like polypeptide (I).

XX The polypeptide is useful for identifying a modulator of (I), by

CC contacting (I) (I) or the nucleotide (II) encoding it, are useful as

CC models for the development of human therapeutics, for identifying

CC therapeutic proteins, and as targets for development of human therapeutic

CC agents. (I) is useful as a reagent in assays designed to quantitatively

CC determine levels of the protein in biological fluids, as markers for

CC tissues in which the corresponding protein is preferentially expressed,

CC in drug screening assays, to identify compounds that modulate human Ras-

CC like protein, or an altered form that causes the specific disease or

CC pathology associated with the protein, to screen a compound for the

CC ability to stimulate or inhibit interaction between the Ras-like protein

CC and a molecule that normally interacts with the Ras-like protein, in

CC pharmacogenomic analysis, and for treating a disorder characterized by

CC altered expression of the human Ras-like protein. (II) is useful to

CC detect the presence of (I) in cells or tissues, to assess abnormal tissue

CC distribution, abnormal subcellular localization or abnormal expression

CC during development or progression of a biological condition, to assess

CC normal in pharmacogenomic analysis, for tissue typing, as diagnostic

CC tools, and for inhibiting a protein function. (III) is for constructing

CC recombinant vectors, host cells and transgenic animals, for expressing

CC antigenic portions of the human Ras-like proteins, for designing

CC ribozymes, for monitoring the effectiveness of modulating compounds on

CC the expression or activity of the human Ras-like gene, and for gene

CC therapy in patients containing cells that are aberrant in human Ras-like

CC protein gene expression. This sequence encodes a novel alternative splice

CC form of human ras-like protein

XX SQ Sequence 88191 BP; 11513 A; 13203 C; 13500 G; 13578 T; 0 U; 36397 Other;

Query Match 3.3%; Score 37.2; DB 8; Length 88191;

Best Local Similarity 59.4%; Pred. No. 9;

Matches 63; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 140 GCGCGGAGCTCCCGAGTTTGAGCCGCAATTTTACATTTATGGCGATGGGTCTAG 199

DB 18903 GCGCGCGGCTCTCTCTTAAAGGCTGAACATGATCTCTTATGGAGAGCCACCTT 18962

QY 200 GCCTTATCTAGCGCTGGGAGGTACATTTGAAGATGTCAC 245

DB 18963 TTCTTTATCCAGTTTCTAGGATGGACACTTGAGGGGCTTCCACC 19008

RESULT 12

ADA71938

ID ADA71938 standard; DNA; 2000 BP.

XX AC ADA71938;

XX 20-NOV-2003 (first entry)

DT Rice gene, SEQ ID 5263.

XX DE

XX Plant; bacterial infection; fungal infection; viral infection; rice;

KW gene; ds.

XX OS Oryza sativa.

XX WO200300898-A1.

XX 03-JAN-2003.

XX 22-JUN-2001; 2001WO-IB001105.

XX 22-JUN-2001; 2001WO-IB001105.

XX (SYGN ) SYNGENTA PARTICIPATIONS AG.

XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;

PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;

XX MPI; 2003-175290/17.

XX Identifying at least one gene involved in plant resistance or response to

PT pathogenic infection for conferring resistance or tolerance to a plant to

PT bacterial, fungal or viral infection by determining or detecting plant

XX gene expression.

XX Claim 27; SEQ ID NO 5263; 899pp; English.

XX The present invention relates to a method (M1) for identifying genes

CC involved in plant resistance or response to pathogenic infection. M1

CC comprises identifying a gene whose expression is significantly altered in

CC the incompatible interaction of plant gene expression relative to

CC the expression of the gene in an uninfected plant, in a mutant plant that

CC does not express a gene associated with response to pathogenic infection,

CC or in a corresponding incompatible or compatible interaction. (M1) is

CC useful for conferring resistance to resistance or tolerance to a plant to

CC bacterial, fungal or viral infection. The present sequence was used to

CC illustrate the invention.

XX SQ Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

Query Match 3.3%; Score 36.6; DB 8; Length 2000;

Best Local Similarity 10.0%; Pred. No. 1.6;

Matches 41; Conservative 190; Mismatches 179; Indels 1; Gaps 1;

QY 702 GTGAGCAGTCGATGATTCAACAACAATCCGGGGTTCAGCAAGTCGGAATTTTC 761

DB 131 GSGRMGGKSRMSYMMWCYARGCGGCKKKKSGSGWGTCTRGARGGSGWSSGAKYKSGS 190

QY 762 GGCTCATCTACTATGTCGTGGGCTTCAGCGTATTCGCTTAACCGTATTTGAATCAGAC 821

DB 191 MSKRMWMSGSGRSGGRRSAYSYGTSRKYGTWKMTYYSASRCMRYMTTTSYSWACSS 250

QY 822 CTTACAGAGCCAGCGCTCCAGGACCCGTTCCACACGTCAGCCTAAAAAAGAAAAA 881

DB 251 YTWCKSRKSMWMMKMRKMRWSYGYMSYMMWMTAYKKSYSYRWCYMYRGGGWGA 310

QY 882 AACTGTTCATCACAGCCCAT-CTGAACCGTTTCAACAGCCCGCCACGTAATTTGCGCCACC 940

DB 311 TRYWREGYSRMAMMYKMYWYRGYKMGKRGHWAGBMWRMSKACYSYMRWRWMT 370

QY 941 AGCAAGGCGATATCCGTATAGCAGCGCATATAATTTCTGATTCCTGCTGCTGCCGA 1000

DB 371 RRRRWAKSSRTSRKKRKCRCMRKRYKMRGYSRMRSCRRARWKRCSRGRWKMGCGR 430

QY 1001 CAATTTATCTTTGGGAGCGCGGCGGATTCGAGACAGAGCCACACACACAA 1060

DB 431 CMTCRMKSYGMMRWKSWKRMASKYKWMRMYRWRKCKCSRTTWMGKTRGGMGTGRCRY 490

QY 1061 GTGCGCGTGAGAAATCAACAAGCGGTCTTCGCGAGAAGAGAGAGAGAG 1111

DB 491 KKRSGMKRKRRCRRRWGRMYRMRKRYMSARYMTMYRCARKKYSYSAARKAR 541

RESULT 13

ACN91201

ID ACN91201 standard; DNA; 1898 BP.

XX AC ACN91201;

XX 02-DEC-2004 (first entry)

DT Breast cancer related marker, seq id 12351.

XX Cancer; breast; tumour; cytostatic; marker; detection; therapy; ds.

XX Homo sapiens.

XX US2003099974-A1.

XX 29-MAY-2003.

XX	18-JUL-2002; 2002US-00198846.	PA	(SYGN ) SYNGENTA PARTICIPATIONS AG.
PF		XX	Zhu T, Cheng W, Briggs S, Cooper B, Goff SA, Moughamer T;
XX	18-JUL-2001; 2001US-0306220P.	PI	Glazebrook J, Katagiri F, Kreps J, Provart N, Ricke D;
PR		XX	WPI; 2003-229341/22.
XX	(MILL-) MILLENNIUM PHARM INC.	DR	
XX	Lillie J, Xu Y, Wang Y, Steinmann K;	XX	
XX	WPI; 2003-787014/74.	XX	
XX	Novel isolated polypeptide associated with breast cancer, useful for	PT	New plant genes encoding polypeptides having an activity involved in or
PT	detecting presence of polypeptide in sample, as a marker for breast	PT	associated with the synthesis, metabolism or degradation of carbohydrates
PT	cancer.	PT	in the plant grain useful in generating plants having improved
PT		PT	nutritional properties.
XX	Disclosure; SEQ ID NO 12351; 36pp; English.	PS	
XX		XX	Disclosure; SEQ ID NO 719; 130pp; English.
XX	The invention relates to an isolated polypeptide (I) associated with a	CC	This invention, in the area of plant biotechnology, relates to novel
CC	breast cancer which is encoded by a nucleic acid molecule comprising a	CC	polynucleotides comprising a nucleotide sequence encoding a protein which
CC	nucleotide sequence (S1). Further disclosed is an antibody that binds to	CC	is involved in or associated with the synthesis, metabolism or
CC	the polypeptide of the invention. The activity of the polypeptide of the	CC	degradation of carbohydrates in the plant grain and the expression of
CC	invention may be described as cytostatic. The antibody is useful for	CC	which is up-regulated during grain filling. The plant is selected from
CC	detecting the presence of (I) in a sample. Nucleic acid molecules of the	CC	corn, tomato, banana, canola, cotton, peanut, sorghum, tobacco,
CC	invention are useful in the detection of breast tumours. (I) is useful as	CC	sugarbeet, wheat, and rice. The invention may be useful for the
CC	a marker for breast cancer and in breast cancer therapy. Sequences given	CC	improvement of protein, oil, starch, fibre and moisture content of the
CC	in records ACN7851-ACN92934 represent nucleic acid markers associated	CC	cereal grains. In addition, carbohydrate levels may be modified to a more
CC	with breast cancer. Note: The sequence listing does not form part of the	CC	desirable level using the present invention. The present sequence is a
CC	specification but may be obtained in electronic format from the USPTO web	CC	sequence of a rice gene promoter. Note: The sequence data for this
CC	site at seqdata.uspto.gov/sequence.html?DocID=2003009974	CC	patent did not form part of the printed specification, but was obtained
CC		CC	in electronic format directly from WIPO at
XX		CC	ftp.wipo.int/pub/publishedpct_sequences.
XX	Sequence 1898 BP; 448 A; 416 C; 492 G; 538 T; 0 U; 4 Other;	SQ	Sequence 2000 BP; 480 A; 520 C; 441 G; 559 T; 0 U; 0 Other;
Query Match 3.3%; Score 36.2; DB 11; Length 1898;			
Best Local Similarity 57.5%; Pred. No. 2;			
Matches 65; Conservative 0; Mismatches 48; Indels 0; Gaps 0;			
QY	963 GCGAGCGCATAAATTCGTATTCCTGCGCTGCGCGGACAAATTTATCTTTGGGAGCGG 1022	QY	344 CCCTTTTAGAGCAGTATATAAGTCTCTAGTCAGCTGGCTATAGATGTTCACATCAGCA 403
Db	32 GGGAGTCCACAGAAAATGGGGCTCACTGACTGTCTGTCAGGTATCTCAGTGGGAGGGA 91	Db	112 CAGCTTAGGGCAAGTACAATAGGACTATCACTACTATTAGCCGTCAGGTCTATCA 53
QY	1023 GCCGGGATGGAGACAGCCCAACAGCAACAAAGTCGCGTGAGAAAT 1075	QY	404 AATCCTTAACTGGAGGAGAAAGAAAGTAGGAG 436
Db	92 ATGGAGTGGGAAATGGGAGTACTCAAGGCCACAGACAGCGCCGACATAGAACT 144	Db	52 CAATC-TGAGCTGGAGGACAGAGAAAGGAAGAG 21
RESULT 14			
ADC08414/c			
ID	ADC08414 standard; DNA; 2000 BP.	ID	ADH22268 standard; DNA; 5955 BP.
XX		XX	
AC	ADC08414;	AC	ADH22268;
XX		XX	
DT	18-DEC-2003 (first entry)	DT	11-MAR-2004 (first entry)
XX		XX	
DE	Rice DNA sequence Seq ID719 related to grain filling.	DE	Rice PONG_LIKE_2 DNA sequence SeqID.
XX		XX	
XX	plant biotechnology; carbohydrate synthesis; carbohydrate metabolism;	XX	transposable element; rice; mping/Pong; transposase; transgenic; plant;
KW	carbohydrate degradation; carbohydrate; plant grain; grain filling; corn;	KW	Pong-like transposase element; PTE; ds.
KW	tomato; banana; canola; cotton; peanut; sorghum; tobacco; sugarbeet;	XX	
KW	wheat; rice; protein; oil; starch; fibre; moisture content; cereal grain;	OS	Oryza sativa.
KW	gene; ds; plant.	XX	
XX		XX	WO2003093451-A2.
OS	Oryza sativa.	XX	
XX		PD	13-NOV-2003.
XX		XX	
PN	WO2003000905-A2.	XX	
XX		PF	16-JAN-2003; 2003WO-US001573.
XX		XX	
PD	03-JAN-2003.	PR	01-MAY-2002; 2002US-0377409P.
XX		XX	
PF	21-JUN-2002; 2002WO-IB002450.	PA	(UYGE-) UNIV GEORGIA RES FOUND INC.
XX		XX	
XX	22-JUN-2001; 2001US-0300112P.	PI	Wessler SR, Jiang N, Bao Z, Zhang X, Eddy SR;
PR	26-SEP-2001; 2001US-0325277P.	XX	
PR	20-DEC-2001; 2001US-0342327P.	DR	WPI; 2004-022658/02.
XX		XX	

PT New transposable element comprising at least a portion of a nucleic acid  
PT comprising two terminal inverted repeat nucleic acid sequences, useful in  
XX designing oligonucleotide primers.

PS Claim 6; SEQ ID NO 17; 202pp; English.

XX  
CC This invention relates to novel isolated transposable elements from rice  
CC and the encoded proteins thereof. Specifically, it refers to the  
CC polynucleotides of the mPing/Pong family of transposable elements that  
CC each comprise two terminal inverted repeat nucleic acid sequences, where  
CC the transposable element is actively transposing. Furthermore, these  
CC polynucleotides encode proteins that can function as transposases or  
CC proteins that interact with transposases to modulate transposition of  
CC members of the mPing/Pong genus. The present invention describes  
CC transgenic plants transformed by Pong-like transposase elements (PTBs),  
CC as well as recombinant expression vectors and appropriate antibodies. As  
CC such, the transposable element is useful for designing oligonucleotide  
CC primers that can be used, in turn, for the isolation of related members  
CC of the mPing/Pong family of transposable elements, and for detecting  
CC transpositions of the transposable element. This polynucleotide is the  
CC DNA sequence of a member of the rice mPing/Pong family, in an  
CC exemplification of the invention.

XX  
SQ Sequence 5955 BP; 1788 A; 1252 C; 1123 G; 1792 T; 0 U; 0 Other;

Query Match 3.3%; Score 36.2; DB 12; Length 5955;  
Best Local Similarity 59.0%; Pred. No. 4;  
Matches 62; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 366 GTCTAGTCAGCTGGCTATAGATGTTCCACATCAGCAATCCTTAACTGGAGGAGAAA 425  
|||||  
DB 5915 GTCCAGGTCAATCAATCTCTGAGCTGGAGGACAGAGAAAGGAGAGAGCTGGAGGACAGA 5856  
|||||

QY 426 GAAAGTAGGAGTGAGAGAGGCGCTCGCGCTTCGTCATCGCTAGC 470  
|||||

DB 5855 GAAAGGAGAGAGAGAGAGAAAGCGCGCATGTTTATTCGACAGC 5811  
|||||

Search completed: August 30, 2005, 13:50:38  
Job time : 702 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 30, 2005, 13:00:09 ; Search time 228 Seconds  
(without alignments)  
7973.261 Million cell updates/sec

**Title:** US-09-979-549-2

**Perfect score:**

Sequence: 1 gaagtcagaaggccgttcag.....ccgagaagagagagagag 1111

Scoring table: IDENTITY NUC

Gapop 10.0 ; Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Minimum DB seq length:	0
Maximum DB seq length:	2000000000

Post-processing: Minimum Match 0%

Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA: \*

```

1: /cgn2_6/prodata/1/ina/5A COMB.seq:*
2: /cgn2_6/prodata/1/ina/5B COMB.seq:*
3: /cgn2_6/prodata/1/ina/6A COMB.seq:*
4: /cgn2_6/prodata/1/ina/6B COMB.seq:*
5: /cgn2_6/prodata/1/ina/PCTUS COMB.seq:*
6: /cgn2_6/prodata/1/ina/backfiles1.seq:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
C	1	39.8	3.6	38503	4	US-09-949-016-17257
	2	39.8	3.6	39715	4	Sequence 17257, A
	3	39.2	3.5	601	4	Sequence 12454, A
	4	38	3.4	27465	4	Sequence 170604, A
	5	37.6	3.4	832	4	Sequence 15561, A
	6	37.2	3.3	601	4	Sequence 2813, Ap
	7	37.2	3.3	601	4	Sequence 31783, A
	8	37	3.3	289	3	Sequence 193904, A
	9	37	3.3	289	3	Sequence 17, Appl
	10	36.2	3.3	53915	4	Sequence 17, Appl
C	11	36.2	3.3	53915	4	Sequence 12221, A
	12	36.2	3.3	53924	4	Sequence 12730, A
	13	36.2	3.3	53924	4	Sequence 12947, A
	14	35.8	3.2	460	4	Sequence 32264, A
	15	35.2	3.2	1141	4	Sequence 32264, A
	16	35	3.2	27968	4	Sequence 22, Appl
	17	35	3.2	27968	4	Sequence 15191, A
	18	35	3.2	108310	4	Sequence 15192, A
	19	34.8	3.1	541	4	Sequence 16366, A
	20	34.2	3.1	11740	4	Sequence 10710, A
C	21	33.8	3.0	56832	3	Sequence 13670, A
	22	33	3.0	168575	3	Sequence 12376, A
	23	32.6	2.9	601	4	Sequence 1, Appl
	24	32.6	2.9	12603	4	Sequence 57820, A
	25	32.6	2.9	20495	4	Sequence 17096, A
	26	32.6	2.9	42988	4	Sequence 17198, A
	27	32.6	2.9	134292	4	Sequence 128, App
	28	32.6	2.9	134292	4	Sequence 12158, A
	29	32.6	2.9	134292	4	Sequence 12158, A
	30	32.6	2.9	134292	4	Sequence 12158, A

## RESULT 1

```

US-09-949-016-17257
: Sequence 17257, Application US/09949016
: Patent No. 6812339
: GENERAL INFORMATION:
: APPLICANT: VENTER, J. Craig et al.
: TITLE OF INVENTION: POLYMORPHISMS IN KNOWN
: TITLE OF INVENTION: WITH HUMAN DISEASE,
: FILE REFERENCE: CLO01307
: CURRENT APPLICATION NUMBER: US/09/949,016
: CURRENT FILING DATE: 2000-04-14
: PRIOR APPLICATION NUMBER: 60/241,755
: PRIOR FILING DATE: 2000-10-20
: PRIOR APPLICATION NUMBER: 60/237,768
: PRIOR FILING DATE: 2000-10-03
: PRIOR APPLICATION NUMBER: 60/231,498
: PRIOR FILING DATE: 2000-09-08
: NUMBER OF SEQ ID NOS: 207012
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 17257
: LENGTH: 38503
: TYPE: DNA
: ORGANISM: Human
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)...(38503)
: OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17257

```

Query Match	3.6%	Score 39.8;	DB 4;	Length 38503;
Best Local Similarity	50.8%;	Pred. No. 0.14;		
Matches 95;	Conservative	0;	Mismatches 92;	Indels 0;
				Gaps 0;

Qy	133	TGGGTGGCGCGAGCTCCCGAGTTTGAGCCGCAATTTTTTTACATTTTATGGCGATG	192
Db	20933	TAGTGTGAGCCACTGCACCCGCGCTTGAAGAGGCAATTTTTTTTTTTTGGAGATG	20992
Qy	193	GCCTCAGGCGTTTATCTAGCGCTCGGAGGGGTACATTTGAAGATGTGCACCACTCCA	252
Db	20993	GAGTTTCGCTCTGTGTGTGGTGGCCAGCTGGAGTGCAAATGGCACCATCTCAGTTTACCGCA	21052

Qy 253 AACCGAACCCCTGTATCTGAGCATGCCCTCATGCTCTCCTTCATGCCTCCCTTTGGTG 312  
| | | | | | | | | | | | | | | | | | | | |  
Db 21053 ACCTCTCCCTCCGGGTTCCAGCGATTCTCTGCTCACTCCGAGTAGTGGGATTA 21112  
| | | | | | | | | | | | | | | | | | | | |  
Qy 313 AGGTCAAT 319

Db 21113 CAGGCAT 21119

## ALIGNMENTS

```

RESULT 1
US-09-949-016-17257
; Sequence 17257, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CU601307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17257
; LENGTH: 38503

```

```
RESULT 2
US-09-949-016-12454
; Sequence 12454, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12454
; LENGTH: 39715
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(39715)
; OTHER INFORMATION: n = A, T, C or G
US-09-949-016-12454

Query Match      3.6%; Score 39.8; DB 4; Length 39715;
Best Local Similarity 50.8%; Pred. No. 0.14;
Matches 95; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

Qy 133 TGGTGGGGCGCGAGCTCCCGAGTTTGAGCCGCAATTTTACATTTTATGGCGATG 192
Db 20933 TAGGTGTGAGCACTGCACCGCGCTTAGAAGGCAATTTTTTTTTTTTTTTTGAGATG 20992

Qy 193 GCGTCAGCGTTTATCTAGGCGTCTGGAGGGTACATTGAAGATGCGCCACCACTCCA 252
Db 20993 GAGTTTCGCTTGTGTGTCAGGCTGGAGTGCATGCAATCTCAGTTCACCGCA 21052

Qy 253 AACCGACAACTGTATCTAGCATGCTCATGCTCTCCCTTCATGCTCCCTTTGGGTG 312
Db 21053 ACCTTCCTCCCGGTTCCAGCATTTCTCTGCTCAGCTCCCGAGTAGCTGGGATTA 21112

Qy 313 AGTTCAT 319
Db 21113 CAGGCAT 21119

RESULT 3
US-09-949-016-170604
; Sequence 170604, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 170604
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
```

```
US-09-949-016-170604

Query Match      3.5%; Score 39.2; DB 4; Length 601;
Best Local Similarity 55.2%; Pred. No. 0.014;
Matches 74; Conservative 1; Mismatches 59; Indels 0; Gaps 0;

Qy 179 ATTTTATGGCGATGGCGTCAAGCGTTTATCTAGGCGTCTGGAGGGGTACATTTGAAGATG 238
Db 222 ATCTGGTGGGCTGGTGACAGGCCTGTCTGCAGAGGACTTGTCTAAGAAAGATGAAGATG 281

Qy 239 TGCCACCAACTCCAAACCGACAACCTGTATCTGAGCATGCTCATGCTCTCCTTCATG 298
Db 282 CAACACCCCAACTCTCCRCCTGCTAGATGGGAGCATGGAGCAACCAACCTTTGCTG 341

Qy 299 CCTCCCTTTGGGTG 312
Db 342 CCTGCTGGGGTG 355

RESULT 4
US-09-949-016-16561
; Sequence 16561, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16561
; LENGTH: 27465
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16561

Query Match      3.4%; Score 38; DB 4; Length 27465;
Best Local Similarity 55.2%; Pred. No. 0.46;
Matches 74; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Qy 179 ATTTTATGGCGATGGCGTCAAGCGTTTATCTAGGCGTCTGGAGGGGTACATTTGAAGATG 238
Db 17411 ATCTGGTGGGCTGGTGACAGGCCTGTCTGCAGAGGACTTGTCTAAGAAAGATGAAGATG 17470

Qy 239 TGCCACCAACTCCAAACCGACAACCTGTATCTGAGCATGCTCATGCTCTCCTTCATG 298
Db 17471 CAACACCCCAACTCTCCCGCTGGCTAGATGGAGCATGGAGCAACCAACCTTTGCTG 17530

Qy 299 CCTCCCTTTGGGTG 312
Db 17531 CCTGCTGGGGTG 17544

RESULT 5
US-09-621-976-2813/c
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
```



```

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc feature
; LOCATION: (1)..(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

Query Match          3.3%; Score 37; DB 3; Length 289;
Best Local Similarity 10.6%; Pred. No. 0.049;
Matches 19; Conservative 75; Mismatches 86; Indels 0; Gaps 0;

Qy 170 TTTTTCATTTATGGCGATGGCGTTCAGGGCTTTATCTAGGCGTCTGGGAGGTACAT 229
    ||||| ||||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db 274 TTTTTCATTTATGGCGATGGCGTTCAGGGCTTTATCTAGGCGTCTGGGAGGTACAT 229
    ||||| ||||| : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 230 TTGAAGATGCCACCAACTCCAAACCGACACCCCTGTATCTGAGCATGCCCTCATGCCCTC 289
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 214 NYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNY 155
Qy 290 TCCTTCATGCTCCCTTTGGGTGAGGTGATGTCCTGGCGGCGAGTGGCTTCCCGTTT 349
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 154 NYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNY 95

RESULT 9
US-09-244-796-17/c
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Sztostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc_feature
; LOCATION: (1)..(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match          3.3%; Score 37; DB 3; Length 289;
Best Local Similarity 10.6%; Pred. No. 0.049;
Matches 19; Conservative 75; Mismatches 86; Indels 0; Gaps 0;

Qy 170 TTTTTCATTTATGGCGATGGCGTTCAGGGCTTTATCTAGGCGTCTGGGAGGTACAT 229
    ||||| ||||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db 274 TTTTTCATTTATGGCGATGGCGTTCAGGGCTTTATCTAGGCGTCTGGGAGGTACAT 229
    ||||| ||||| : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 230 TTGAAGATGCCACCAACTCCAAACCGACACCCCTGTATCTGAGCATGCCCTCATGCCCTC 289
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 214 NYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNY 155
Qy 290 TCCTTCATGCTCCCTTTGGGTGAGGTGATGTCCTGGCGGCGAGTGGCTTCCCGTTT 349
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 154 NYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNY 95

RESULT 10
US-09-949-016-12221
; Sequence 12221, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12221
; LENGTH: 53915
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(53915)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12221

Query Match          3.3%; Score 36.2; DB 4; Length 53915;
Best Local Similarity 57.5%; Pred. No. 3;
Matches 65; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 963 GCGAGGCGCATAAATTCCTGATTCCTGCTCGCTCCGCGACAAATTTATCTTTGGGAGGCGG 1022
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 44530 GCGAGTCCACAGAAAATGGGCGCTCACTGACTGCTTGTCCAGTATCTCAGTGGGAGGGA 44589
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1023 GCGGGGATTTGGACAGAGCCCAAGGCAACAAAGTGGCGGTGAGAAAT 1075
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 44590 ATGGAGTGGGAAATGGAGTACTCAAGGCCCAAGACAGGCCCATAGAACT 44642
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
US-09-949-016-12730
; Sequence 12730, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12730
; LENGTH: 53915
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(53915)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12730

```

	Query Match	3.3%	Score 36.2;	DB 4;	Length 53915;
	Best Local Similarity	57.5%;	Pred. No. 3;		
	Matches 65;	Conservative 0;	Mismatches 48;	Indels 0;	Gaps 0;
Qy	963	GCAGGCGCATAAATTCGTATCTCGCTGCTGCCGGACAATTATCTTTGGGGAGGCGG	1022		
Db	44530	GGGAGTCCACAGAAAAATGGGGCCTCACTGACTTGCTTCAGGTATCTCAGTGGGAGGGA	44589		
Qy	1023	CCCGGNATTGGACACAGAGCCCCACAGGCCAACAAAGAATGGCGGTGAGAAAT	1075		
Db	44590	ATGGGAGTGGGAAAATGGAGTACTCTAAGGGCCACAAGACAGCGCCCATAGAACT	44642		

```

RESULT 12
US-09-949-016-12947
; Sequence 12947, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INFECTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12947
; LENGTH: 53924
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(53924)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12947

```

	Query Match	3.3%	Score 36.2;	DB 4;	Length 53924;
	Best Local Similarity	57.5%;	Pred. No. 3;		
	Matches	65;	Conservative 0;	Mismatches 48;	Indels 0; Gaps 0;
Qy	963	GGGAGCGCATAAATTC	TGATTCCTGCCTGCTGCCGGA	CAATTTATCTTTTGGGGAGGCGG	1022
Db	44530	GGGAGTCCACAGAAAT	GGGGCCCTCACTGACTCTTGT	CCAGGTATCTCAGTGGGAGGA	44589
Qy	1023	CCCGGGAATGGGACAGAGCC	CAAGGCCAACAACAAAGTGGCGG	CTGAGAAAT	1075
Db	44590	ATGGGAGTGGGAAATGAGT	TACTCAAGGCCACAAGACAGCGG	CCCATAGAACT	44642

RESULT 13  
US-09-949-016-12948  
; Sequence 12948, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CLO01307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08

```

; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12948
; LENGTH: 53924
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(53924)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12948

```

	Query Match Best Local Similarity Matches	3.3%; 57.5%; 65;	Score 36.2; Pred. No. 3; Conservative 0;	DB 4; Mismatches 48; Indels 0;	Length 53924; Gaps 0;
Qy	963	GGAGGGCAATAATTCTGATTCCTCGCTGCCGGACAATTTATCTTTGGGAGGCGG	1022		
Db	44530	GGAGTCCACAGAAAATGGGCGCTCACTGACTGCTTGTCCAGGTATCTCAGTGGGAGGA	44589		
Qy	1023	GGCGGATTTGGAGACAGAGCCCAAGGCAACAACAAGTGGCGGTGNGAAT	1075		
Db	44590	ATGGAGTGGGAAATGGAGTACTCAAGGCCACAAGACAGCGCGCCATAGAACT	44642		

```

RESULT 14
US-09-513-999C-32264
; Sequence 32264, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 32264
; LENGTH: 460
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 26
; OTHER INFORMATION: s=g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 188
; OTHER INFORMATION: r=a or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 193
; OTHER INFORMATION: m=a or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 194
; OTHER INFORMATION: r=a or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 311
; OTHER INFORMATION: n=a, g, c or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 352
; OTHER INFORMATION: n=a, g, c or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 353

```

```
; OTHER INFORMATION: d=a or g or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 458
; OTHER INFORMATION: y=c or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 459
; OTHER INFORMATION: k=g or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 460
; OTHER INFORMATION: y=c or t
US-09-513-999C-32264
```

```
Query Match 3.2%; Score 35.8; DB 4; Length 460;
Best Local Similarity 56.6%; Pred. No. 0.17; Indels 0; Gaps 0;
Matches 64; Conservative 0; Mismatches 49;

QY 963 GCGAGCGCATAAATTCGTGATTCCTGCTGCTGCGGACAAATTTATCTTTGGGAGGCGG 1022
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
272 GGGAGTCCACAGAAATGGGGCTCACTGACTGTTGTCNAGGTAICTCAGTGGGAGGA 331

QY 1023 GCCGGGATTTGGAGACAGACCCACAGGCAACAAAGTGCCTGAGAAAT 1075
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
332 ATGGGAGTGGAAATGGAGTNDTCAAGGCCACAAGACAGCGCCCATAGAACT 384
```

```
RESULT 15
US-09-806-708B-22
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: {1}..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22
```

```
Query Match 3.2%; Score 35.2; DB 4; Length 1141;
Best Local Similarity 17.5%; Pred. No. 0.51; Indels 0; Gaps 0;
Matches 45; Conservative 79; Mismatches 133;

QY 157 GTTTGAGCGCGCAATTTTACATTTATGGCGATGGCGTTCAGCGGTTTATCTAGCGCTC 216
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
833 GTNNNNNNNNAYAWTKWYITDDRBAYTNNNNNNRMAYYAYADDDYAYMSDTC 892

QY 217 TGGGAGGGTACATTTGAAGATGTGCCACCACTCAACCGACCAACCCCTGTATCTGAGCA 276
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
893 DAWKWDATKNNNATYNRGTAWRTNNNNNNMTKTYBHAAWNNNNNNNGKMCCTAHTWW 952

QY 277 TGCCTCATGCTCTCCCTTCCTCCCTTTGGGAGGTCATGCTGCGCTTGGCGGCGAG 336
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
953 VCKATKTTGGCWMNCCTTCRKIKNCTWYTWTTTTRTWYATRWKTNATGSMTRCNAT 1012

QY 337 TGGCTTCCCGTTTAGAGCAAGTATAAAGTCCTAGTCAGCTGGCTATAAGATCTTCCAC 396
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1013 GWKNNNYTWKTRWTAYRMATRWKAWKVMATGSMWNTNSYARWAYKTRAYKGWYNAC 1072

QY 397 ATCAGCAATCCTTAAA 413
```

```
Db 1073 AWRWRWKATCYMTDNA 1089

Search completed: August 30, 2005, 16:31:26
Job time : 232 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 30, 2005, 13:39:06 ; Search time 859 Seconds  
(without alignments)  
8462.599 Million cell updates/sec

Title: US-09-979-549-2  
Perfect score: 1111  
Sequence: 1 gaagtcagaagccgttcag.....ccgagaagagagagagag 1111

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7331713 seqs, 3271544945 residues  
Total number of hits satisfying chosen parameters: 14663426

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : Published Applications NA:\*
- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
  - 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
  - 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
  - 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
  - 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
  - 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
  - 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
  - 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
  - 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*
  - 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*
  - 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*
  - 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
  - 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*
  - 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*
  - 15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*
  - 16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:\*
  - 17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*
  - 18: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq:\*
  - 19: /cgn2\_6/ptodata/1/pubpna/US10G\_PUBCOMB.seq:\*
  - 20: /cgn2\_6/ptodata/1/pubpna/US10H\_PUBCOMB.seq:\*
  - 21: /cgn2\_6/ptodata/1/pubpna/US10I\_PUBCOMB.seq:\*
  - 22: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
  - 23: /cgn2\_6/ptodata/1/pubpna/US11A\_PUBCOMB.seq:\*
  - 24: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:\*
  - 25: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
  - 26: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	79.6	7.2	629	20	US-10-425-115-15660 Sequence 15660, A
2	47.8	4.3	9538	21	US-10-311-193-9 Sequence 9, Appli
3	45.8	4.1	1976	17	US-10-260-238-5967 Sequence 5967, Ap
4	41	3.7	1999	22	US-10-481-032A-733 Sequence 733, App
5	39.4	3.5	49600	18	US-10-459-262A-4 Sequence 4, Appli
6	39.4	3.5	69300	18	US-10-415-058-6 Sequence 6, Appli
7	38	3.4	24333	11	US-09-984-429-351 Sequence 351, App

c	8	38	3.4	24333	11	US-09-984-429-460	Sequence 460, App
	9	37.2	3.3	88191	9	US-09-799-799-3	Sequence 3, Appli
	10	37	3.3	421	18	US-10-424-599-385	Sequence 385, App
	11	36.2	3.3	1898	14	US-10-198-846-12351	Sequence 12351, A
c	12	36.2	3.3	2000	22	US-10-481-032A-719	Sequence 719, App
c	13	36.2	3.3	5955	18	US-10-346-198-22	Sequence 22, Appl
	14	35.2	3.2	300	20	US-10-425-115-76475	Sequence 76475, A
c	15	35	3.2	704	20	US-10-363-345A-18601	Sequence 18601, A
c	16	35	3.2	704	20	US-10-363-345A-18602	Sequence 18602, A
c	17	35	3.2	704	21	US-10-363-483A-18601	Sequence 18601, A
c	18	35	3.2	704	21	US-10-363-483A-18602	Sequence 18602, A
c	19	34.8	3.1	594	14	US-10-123-155-10	Sequence 10, Appl
c	20	34.8	3.1	594	15	US-10-146-731-10	Sequence 10, Appl
c	21	34.8	3.1	594	15	US-10-140-472-10	Sequence 10, Appl
c	22	34.8	3.1	594	15	US-10-141-761-10	Sequence 10, Appl
c	23	34.8	3.1	594	16	US-10-142-885-10	Sequence 10, Appl
c	24	34.8	3.1	594	16	US-10-158-790-10	Sequence 10, Appl
c	25	34.8	3.1	594	17	US-10-137-871-10	Sequence 10, Appl
c	26	34.8	3.1	594	17	US-10-140-923-10	Sequence 10, Appl
c	27	34.8	3.1	594	17	US-10-141-756-10	Sequence 10, Appl
c	28	34.8	3.1	594	17	US-10-141-759-10	Sequence 10, Appl
c	29	34.8	3.1	594	17	US-10-140-805-10	Sequence 10, Appl
c	30	34.8	3.1	594	17	US-10-140-864-10	Sequence 10, Appl
c	31	34.8	3.1	594	18	US-10-142-426-10	Sequence 10, Appl
c	32	34.8	3.1	777	14	US-10-184-644-348	Sequence 348, App
c	33	34.8	3.1	777	14	US-10-184-634-348	Sequence 348, App
c	34	34.8	3.1	7525	9	US-09-967-552A-72	Sequence 72, Appl
c	35	34.8	3.1	13031	16	US-10-387-894-5	Sequence 5, Appli
c	36	34.6	3.1	837	20	US-10-363-345A-21779	Sequence 21779, A
c	37	34.6	3.1	837	20	US-10-363-345A-21780	Sequence 21780, A
c	38	34.6	3.1	837	21	US-10-363-483A-21779	Sequence 21779, A
c	39	34.6	3.1	837	21	US-10-363-483A-21780	Sequence 21780, A
c	40	34.2	3.1	802	14	US-10-184-644-312	Sequence 312, App
c	41	34.2	3.1	802	14	US-10-184-634-312	Sequence 312, App
c	42	34.2	3.1	8649	17	US-10-221-613-117	Sequence 117, App
c	43	34.2	3.1	8649	18	US-10-221-714A-51	Sequence 51, Appl
c	44	34.2	3.1	118931	13	US-10-087-192-1108	Sequence 1108, Ap
c	45	34	3.1	561	21	US-10-487-901-191	Sequence 191, App

ALIGNMENTS

RESULT 1

US-10-425-115-15660  
; Sequence 15660, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plancs  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 15660  
; LENGTH: 629  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_114281C.1  
US-10-425-115-15660

Query Match	7.2%	Score	79.6;	DB	20;	Length	629;
Best Local Similarity	92.3%;	Pred. No.	3.1e-15;				
Matches	96;	Conservative	0;	Mismatches	4;	Indels	1;
Qy	1008	TCCTTTGGGAGCGCGCGGATTCGAGACAGAGCCCAAGCCACACAAAGTCGCGG	1067				
Db	1	TATTTGGGAGCGCGCGGATTCGAGACAGAGCCCAAGCCACAAAGTCGCGG	60				

```
QY 1068 TGAGMATCAACAGCGTCTTCCGAGAGAGAGAGAG 1111
Db 61 TGAGA----AACAGCGTCTTCCGAGAGAGAGAGAG 100

RESULT 2
US-10-311-193-9
; Sequence 9, Application US/10311193
; Publication No. US2005009845A1
; GENERAL INFORMATION:
; APPLICANT: SPANGENBERG, German Carlos
; APPLICANT: LIDGETT, Angela Jane
; APPLICANT: JOHNSON, Xenie Angela
; APPLICANT: TERDICH, Katherine
; TITLE OF INVENTION: FRUCTOSYL TRANSFERASE HOMOLOGUES FROM RYEGRASS (LOLIUM) AND
; FILE REFERENCE: 4491-4001
; CURRENT FILING DATE: 2002-12-13
; PRIOR APPLICATION NUMBER: P08155
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: PCT/AU01/00705
; PRIOR FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 9538
; TYPE: DNA
; ORGANISM: Lolium perenne
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3698)..(3698)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4669)..(4669)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (5546)..(5546)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (5611)..(5611)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (5615)..(5615)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (6033)..(6033)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (6221)..(6221)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (6404)..(6404)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (6502)..(6502)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (6528)..(6528)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (6533)..(6533)

; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (6536)..(6536)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (6619)..(6619)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (6627)..(6627)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (6632)..(6632)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (6647)..(6647)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (6675)..(6675)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (6696)..(6696)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (6735)..(6735)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (6749)..(6749)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (7516)..(7516)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9095)..(9095)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9121)..(9121)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9260)..(9260)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9471)..(9471)
; OTHER INFORMATION: n is a, c, g, or t
; US-10-311-193-9

Query Match 4.3%; Score 47.8; DB 21; Length 9538;
Best Local Similarity 62.9%; Pred. No. 0.00057;
Matches 90; Conservative 0; Mismatches 52; Indels 1; Gaps 1;

QY 350 AGAGCAAGTATAATAGTCTTAGTCAGCTGGCTATAGATGTTCCACATCAGCAAAATCCT 409
Db 7989 AGAGCAAGTACATAGATCTAGTCAGCTGGCTACAGGATTAAATAATATATTTGTGT 8048
QY 410 TAAACTGGAGGAGAAAGTAGGAGTGAAGGGCGTCGGCGCTTCGTCATCGCTAG 469
Db 8049 CTAGTTGGAGGAGATAGGAGGAGAGA-ATGTGAGTATGCTCTTTATCGAAGAGCTAG 8107
QY 470 CGATAGCACAAAGCTCCCATGAA 492
```



```
Db      8108 CTCTAGCAGCTGCTCTCTAGGCAA 8130

RESULT 3
US-10-260-238-5967/c
; Sequence 5967, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 5967
; LENGTH: 1976
; TYPE: DNA
; ORGANISM: Oryza sativa
US-10-260-238-5967

Query Match      4.1%; Score 45.8; DB 17; Length 1976;
Best Local Similarity 58.8%; Pred. No. 0.0012;
Matches 97; Conservative 0; Mismatches 67; Indels 1; Gaps 1;

QY      348 TTAGCAGCAGTAATAAGTCCTAGTCAGCTGGCTATAGATGTTCCACATCAGCAATC 407
         |||||
Db      696 TAAGAGCAAGTATAGTAATGTACAGTCAGCTGGCGAAATCAACCGTGGCTCATCCAAAT 637
         |||||

QY      408 CTTAAACTGGAGGAGAAAGAACTAGGAGTGAGAGGGCGCTCGCGCTTCGTCAATCGCT 467
         |||||
Db      636 C-CAACGTGGAGGCGGAGAACCAAGAGAGAGAAATGGCGAGTCGCGAGTCGCC 578
         |||||

QY      468 AGCGATAGCAACAGCTCCCATGGAATCGAGCCCAACATGCNAACCCG 512
         |||||
Db      577 GGCTCGAAGCGTGTACCCGAGGCAAAACGCCCGCTCCCAACTCG 533
         |||||

RESULT 4
US-10-481-032A-733/c
; Sequence 733, Application US/10481032A
; Publication No. US20050177901A1
; GENERAL INFORMATION:
; APPLICANT: Zhu, Tong
; APPLICANT: Cheng, Wenqiong
; APPLICANT: Briggs, Steven
; APPLICANT: Cooper, Bret
; APPLICANT: Goff, Stephen A.
; APPLICANT: Moughamer, Todd
; APPLICANT: Glazebrook, Jane
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES
; FILE REFERENCE: 60148USPT
; CURRENT APPLICATION NUMBER: US/10/481,032A
; CURRENT FILING DATE: 2003-12-16

; PRIOR APPLICATION NUMBER: US 60/300,112
; PRIOR FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/342,327
; PRIOR FILING DATE: 2001-12-20
; PRIOR APPLICATION NUMBER: PCT/IB02/02450
; PRIOR FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 1201
; SOFTWARE: PatentIn Ver. 2.2
; SEQ ID NO 733
; LENGTH: 1999
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1730)..(1730)
; OTHER INFORMATION: n = any nucleotide
US-10-481-032A-733

Query Match      3.7%; Score 41; DB 22; Length 1999;
Best Local Similarity 72.6%; Pred. No. 0.047;
Matches 53; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY      628 TATAGCTGACCTGGCAGTCTATAGAGCGGCGGCTCTTCTATTAGCTTGTCTT 687
         |||||
Db      282 TATAGATGACATGGTAGTAGTATTCAGCCAGCGCGCTTAAAGTATTAGCCTTGTCTT 223
         |||||

QY      688 ATGGCTACATCTG 700
         |||||
Db      222 AGGAGGGATGTG 210
         |||||

RESULT 5
US-10-459-262A-4
; Sequence 4, Application US/10459262A
; Publication No. US20040083501A1
; GENERAL INFORMATION:
; APPLICANT: Leong, Sally A.
; APPLICANT: Chauhan, Rajinder S.
; APPLICANT: Durfee, Timothy J.
; APPLICANT: Farman, Mark L.
; TITLE OF INVENTION: Plant Genes That Confer Resistance to Strains of
; TITLE OF INVENTION: Magnaporthe Grisea Having AVR1 C039 Cultivar
; FILE REFERENCE: 0141.03
; CURRENT APPLICATION NUMBER: US/10/459,262A
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: 10/415,058
; PRIOR FILING DATE: 2003-04-11
; PRIOR APPLICATION NUMBER: PCT/US01/46331
; PRIOR FILING DATE: 2001-10-19
; PRIOR APPLICATION NUMBER: PCT WO 02/34927
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: 60/242,313
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/303,897
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 49600
; TYPE: DNA
; ORGANISM: Magnaporthe grisea
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(49600)
; OTHER INFORMATION: Continuation of Sequence ID 1, representing bases
; OTHER INFORMATION: 148,881 through 198,480
US-10-459-262A-4

Query Match      3.5%; Score 39.4; DB 18; Length 49600;
Best Local Similarity 65.2%; Pred. No. 0.9;
```

Matches 58; Conservative 0; Mismatches 31; Indels 0; Gaps 0;  
QY 601 CAGCCAGTTTATTATATAACAGGCTATATAGCTGACCTGGCAGTGTCTATAGAGCGGCA 660  
||||| |||||||  
Db 48694 CAGCGGCTTATTATATAAGTGAAGGCTTTAGGTGATGTCTACTAGTATACGCCCACT 48753  
||||| |||||||  
QY 661 GCCGGCTCTTCTATTAGCTTTGCTCTTAT 689  
||||| |||||||  
Db 48754 GCTGGCGTCTTCTATTATCTCTTGCTCTTAT 48782  
||||| |||||||

RESULT 6  
US-10-415-058-6  
; Sequence 6, Application US/10415058  
; Publication No. US20040060081A1  
; GENERAL INFORMATION:  
; APPLICANT: Wisconsin Alumni Research Foundation  
; APPLICANT: United States Department Of Agriculture  
; APPLICANT: Leong, Sally A.  
; APPLICANT: Farman, Mark L.  
; APPLICANT: Chauhan, Rajinder  
; APPLICANT: Durfee, Timothy J.  
; TITLE OF INVENTION: Plant Gene That Confers Resistance To Strains Of Magnaporthe Grisea  
; FILE REFERENCE: Having AVR C039 Cultivar Specificity Gene  
; CURRENT APPLICATION NUMBER: US/10/415,058  
; PRIOR FILING DATE: 2003-04-11  
; PRIOR APPLICATION NUMBER: USN 60/242,313  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: USN 60/303,897  
; PRIOR FILING DATE: 2001-07-09  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6  
; LENGTH: 69300  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (26297)..(26395)  
; OTHER INFORMATION: N is any nucleotide  
US-10-415-058-6

Query Match 3.5%; Score 39.4; DB 18; Length 69300;  
Best Local Similarity 65.2%; Pred. No. 1.1;  
Matches 58; Conservative 0; Mismatches 31; Indels 0; Gaps 0;  
QY 601 CAGCCAGTTTATTATATAACAGGCTATATAGCTGACCTGGCAGTGTCTATAGAGCGGCA 660  
||||| |||||||  
Db 35667 CAGCGGCTTATTATATAAGTGAAGGCTTTAGGTGATGTCTACTAGTATACGCCCACT 35726  
||||| |||||||  
QY 661 GCCGGCTCTTCTATTAGCTTTGCTCTTAT 689  
||||| |||||||  
Db 35727 GCTGGCGTCTTCTATTATCTCTTGCTCTTAT 35755  
||||| |||||||

RESULT 7  
US-09-984-429-351/c  
; Sequence 351, Application US/09984429  
; Publication No. US20040010132A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 53 Human Secreted Proteins  
; FILE REFERENCE: P2018P2  
; CURRENT APPLICATION NUMBER: US/09/984,429  
; CURRENT FILING DATE: 2001-10-30  
; PRIOR APPLICATION NUMBER: 60/244,591  
; PRIOR FILING DATE: 2000-11-01  
; PRIOR APPLICATION NUMBER: 09/288,143  
; PRIOR FILING DATE: 1999-04-08  
; PRIOR APPLICATION NUMBER: PCT/US98/21142  
; PRIOR FILING DATE: 1998-10-08  
; PRIOR APPLICATION NUMBER: 60/061,463

; PRIOR FILING DATE: 1997-10-09  
; PRIOR APPLICATION NUMBER: 60/061,529  
; PRIOR FILING DATE: 1997-10-09  
; PRIOR APPLICATION NUMBER: 60/071,498  
; PRIOR FILING DATE: 1997-10-09  
; PRIOR APPLICATION NUMBER: 60/061,527  
; PRIOR FILING DATE: 1997-10-09  
; PRIOR APPLICATION NUMBER: 60/061,536  
; PRIOR FILING DATE: 1997-10-09  
; PRIOR APPLICATION NUMBER: 60/061,532  
; PRIOR FILING DATE: 1997-10-09  
; NUMBER OF SEQ ID NOS: 727  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 351  
; LENGTH: 24333  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-984-429-351

Query Match 3.4%; Score 38; DB 11; Length 24333;  
Best Local Similarity 55.2%; Pred. No. 1.8;  
Matches 74; Conservative 0; Mismatches 60; Indels 0; Gaps 0;  
QY 179 ATTTTATGGGATGGCTCAGCGTTTATCTAGGCGTCTGGGAGGTACATTTGAAGATG 238  
||||| |||||||  
Db 8903 ATCTGGTGGGCTGGTGACAGGCCTGTCTGCAGAGGACTTGTCTAAGAAAGATGAAGATG 8844  
||||| |||||||  
QY 239 TGCCACCAACTCCAAACCGACAACCTGTATCTTGAGCATGCGCTCATGCTCTCTCTTCATG 298  
||||| |||||||  
Db 8843 CAACACCCAACTCTCCCGCTGGCTAGATGGGAGCATGGACCAACCACTTTGCTG 8784  
||||| |||||||  
QY 299 CTCTCCCTTTGGGTG 312  
||||| |||||||  
Db 8783 CTCTCTGGGTG 8770  
||||| |||||||

RESULT 8  
US-09-984-429-460/c  
; Sequence 460, Application US/09984429  
; Publication No. US20040010132A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 53 Human Secreted Proteins  
; FILE REFERENCE: P2018P2  
; CURRENT APPLICATION NUMBER: US/09/984,429  
; CURRENT FILING DATE: 2001-10-30  
; PRIOR APPLICATION NUMBER: 60/244,591  
; PRIOR FILING DATE: 2000-11-01  
; PRIOR APPLICATION NUMBER: 09/288,143  
; PRIOR FILING DATE: 1999-04-08  
; PRIOR APPLICATION NUMBER: PCT/US98/21142  
; PRIOR FILING DATE: 1998-10-08  
; PRIOR APPLICATION NUMBER: 60/061,463  
; PRIOR FILING DATE: 1997-10-09  
; PRIOR APPLICATION NUMBER: 60/061,529  
; PRIOR FILING DATE: 1997-10-09  
; PRIOR APPLICATION NUMBER: 60/071,498  
; PRIOR FILING DATE: 1997-10-09  
; PRIOR APPLICATION NUMBER: 60/061,527  
; PRIOR FILING DATE: 1997-10-09  
; PRIOR APPLICATION NUMBER: 60/061,536  
; PRIOR FILING DATE: 1997-10-09  
; PRIOR APPLICATION NUMBER: 60/061,532  
; NUMBER OF SEQ ID NOS: 727  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 460  
; LENGTH: 24333  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-984-429-460

Query Match 3.4%; Score 38; DB 11; Length 24333;

```
Best Local Similarity 55.2%; Pred. No. 1.8;
Matches 74; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 179 ATTTTATGGCGATGGCGTTCAGGCGTTTATCTAGGCGTCTGGGAGGCTACATTTGAAGATG 238
Db 8903 ATCTGTGGGCTGGTGACAGGCTGTCTGCAGAGACTGTCTAGAAGAAGATGAAGATG 8844

QY 239 TGCACAACTCCAAACCGCAACCTGTATCTGTAGCATGCCCTTCATGCTCTCTTCATG 298
Db 8843 CAACACCAAACTCTCCCGCTGGCTAGATGGGAGCATGGAGCAAAACCAACCTTTGCTG 8784

QY 299 CTCTCCCTTTGGGTG 312
Db 8783 CTCTGCTGTGGGTG 8770

RESULT 9
US-09-799-799-3
; Sequence 3, Application US/09799799
; Patent No. US20020132291A1
; GENERAL INFORMATION:
; APPLICANT: YE, Jane et al.
; TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL001157
; CURRENT APPLICATION NUMBER: US/09/799,799
; CURRENT FILING DATE: 2001-03-07
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 88191
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(88191)
; OTHER INFORMATION: n = A,T,C or G
US-09-799-799-3

Query Match 3.3%; Score 37.2; DB 9; Length 88191;
Best Local Similarity 59.4%; Pred. No. 6.7;
Matches 63; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 140 GCCTGCGAGCTCCCGAGTTTGACGCGCAATTTTTTACATTTTATGGCGATGGCGTCAG 199
Db 18903 GCCTCGGCTCTCTCTGTTAAGGCTGAACATGATCTCTTCTATGGAGGCCACCTT 18962

QY 200 GCCTTATCTAGCGCTGCGGAGGGTACATTTGAAGATGTCACC 245
Db 18963 TTCTTTATCCAGGTTTCTAGGATGGACACTTGAGGGGCTTCCACC 19008

RESULT 10
US-10-424-599-385
; Sequence 385, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 385
; LENGTH: 421
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:

; NAME/KEY: unsure
; LOCATION: (1)...(421)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_100348C.1
US-10-424-599-385

Query Match 3.3%; Score 37; DB 18; Length 421;
Best Local Similarity 53.9%; Pred. No. 0.44;
Matches 76; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 867 AAAAAAGAAAAAACTGTTCAATCATCACGCCCATCTGAACCGTTCAACAGCCCCAG 926
Db 98 AAAAAAGAAAAAACTGTTCAATCATCACGCCCATCTGAACCGTTCAACAGCCCCAG 926

QY 927 TAAATTTGCGGCACAGCAAGGGCATATCCGTATAGCGGCGCATATAAATTTCTGATTTCT 986
Db 158 TAACTTCCCTTCCATCCAACTAAAATTTTATATAAGGGCCCCCAAAATCAATTCAT 217

QY 987 GCCTGCTGCCGAGCAATTTA 1007
Db 218 TGCCCCCGGTTAAACAACATA 238

RESULT 11
US-10-198-846-12351
; Sequence 12351, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12351
; LENGTH: 1898
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1, 2, 3, 4
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-12351

Query Match 3.3%; Score 36.2; DB 14; Length 1898;
Best Local Similarity 57.5%; Pred. No. 1.8;
Matches 65; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 963 GCGAGCGCATAAATCTGATTCCTCGCTGCGGCAATTTATCTTTGGGAGCGG 1022
Db 32 GGGAGTCCACAGAAATGGGCTCTCACTGACTGCTTGTTCAGGTATCTCAGTGGAGGGA 91

QY 1023 GCCGGGATTTGGAGCAGAGCCCCACAAGGCAACAACAAAGTGGCGTGAGAAAT 1075
Db 92 ATGGGAGTGGGAAATGGAGTACTCAAGGCCACAAGAGCGGCCCATAGAACT 144

RESULT 12
US-10-032A-719/c
; Sequence 719, Application US/10481032A
; Publication No. US20050177901A1
; GENERAL INFORMATION:
; APPLICANT: Zhu, Tong
; APPLICANT: Cheng, Wenqiong
```

```
; APPLICANT: Briggs, Steven
; APPLICANT: Cooper, Bret
; APPLICANT: Goff, Stephen A.
; APPLICANT: Moughamer, Todd
; APPLICANT: Glazebrook, Jane
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicolas
; APPLICANT: Ricke, Darrell
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES
; FILE REFERENCE: 60148USPCT
; CURRENT APPLICATION NUMBER: US/10/481,032A
; CURRENT FILING DATE: 2003-12-16
; PRIOR FILING DATE: US 60/300,112
; PRIOR FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/342,327
; PRIOR FILING DATE: 2001-12-20
; PRIOR APPLICATION NUMBER: PCT/IB02/02450
; PRIOR FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 1201
; SOFTWARE: PatentIn Ver. 2.2
; SEQ ID NO 719
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Oryza sativa
US-10-481-032A-719

Query Match      3.3%; Score 36.2; DB 22; Length 2000;
Best Local Similarity 68.8%; Pred. No. 1.9;
Matches 64; Conservative 0; Mismatches 28; Indels 1; Gaps 1;

Qy 344 CCGTTTATAGCAAGTATAATAAGTCCTAGTCAGCTGGCTATAGATGTTCCACATCAGCA 403
Db 112 CAGCTTAGGCGAGTACATATAGGACTATATCACTACTATTAGCCGTCAGGTCTATCA 53

Qy 404 AATCCTTAAATCGGAGGAGAAAGAAAGTAGGAG 436
Db 52 CAATC-TCAGCTGGAGGACAGAGAAAGGAAGAG 21

RESULT 13
US-10-346-198-22/c
; Sequence 22, Application US/10346198
; Publication No. US20040043485A1
; GENERAL INFORMATION:
; APPLICANT: WESSLER, SUSAN R.
; APPLICANT: JIANG, NING
; APPLICANT: BAO, ZHIRONG
; APPLICANT: ZHANG, XIAOYU
; APPLICANT: EDDY, SEAN R.
; TITLE OF INVENTION: TRANSPOSABLE ELEMENTS IN RICE AND METHODS OF USE
; FILE REFERENCE: 18465-0018
; CURRENT APPLICATION NUMBER: US/10/346,198
; CURRENT FILING DATE: 2003-01-16
; PRIOR FILING DATE: 2003-01-16
; PRIOR APPLICATION NUMBER: 60/337,409
; PRIOR FILING DATE: 2002-05-01
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 5955
; TYPE: DNA
; ORGANISM: Oryza sativa
US-10-346-198-22

Query Match      3.3%; Score 36.2; DB 18; Length 5955;
Best Local Similarity 59.0%; Pred. No. 3.4;
Matches 62; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 366 GTCCAGTTCAGCTGGCTATAGATGTTCCACATCAGCAAAATCCTTAAACTGGAGAGAAA 425
Db 5915 GTCCAGGTCATCAACATCTAGCTGGAGGACAGAGAAAGGAGAGAGCTGGAGGACAGA 5856
```

```
Qy 426 GAAAGTAGCAGTGAAGAAGGCGTCGCGCTTCGTCAATCGCTAGC 470
Db 5855 GAAAGGAGAGAGAGAAAGAGCGGCGCATGTTTAATCGACAGC 5811
```

## RESULT 14

```
US-10-425-115-76475
; Sequence 76475, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 76475
; LENGTH: 300
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_169773C.1
US-10-425-115-76475
```

```
Query Match      3.2%; Score 35.2; DB 20; Length 300;
Best Local Similarity 53.7%; Pred. No. 1.5;
Matches 73; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 920 CCCACGTAAATTTGCGCGCACAGCAAGGGCATATCCGTATAGCGAGCGCATAAATTC 979
Db 2 CGCTCGGAGTGGCGGCTTAACGCAACGGCTCAGCCCTCAGACCAGCAGCAGCATCG 61

Qy 980 GATTCCTGCTGCTGCTCGGACAAATTTATCTTTGGGAGGCGGCGCGGATTTGAGACAG 1039
Db 62 GATTCGTGATCGTACCGCGGTGGACTACCGCTTGTAGTTGGCAGCAGACTTGACAGTGG 121

Qy 1040 AGCCCAACAGGCAACA 1055
Db 122 CTCCAACAGGCAACA 137
```

## RESULT 15

```
US-10-363-345A-18601/c
; Sequence 18601, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Method for determining the degree of methylation of defined
; FILE REFERENCE: E01/1227
; CURRENT APPLICATION NUMBER: US/10/363,345A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18601
; LENGTH: 704
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: CpG-island No: 18601
US-10-363-345A-18601
```

```
Query Match      3.2%; Score 35; DB 20; Length 704;
Best Local Similarity 56.5%; Pred. No. 2.7;
Matches 65; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
```

Qy	815	ATCAGACCCCTCAGAGCCGACGGCTCCAGCGACCCGTTTACACAGTTCAGCCTAAAAAAG	874
Db	133	AACACCGCGTTACTAACGCTTACTCCACCGACGCGTTCAACGAACTAACCGTAAAAACG	74
Qy	875	AAAAAAAACCTGTTCAATCACAGCCCATCTGAACCGTTCAACAGCCCCACGTAA	929
Db	73	AACGCGAACGTATACCAACGCGCCCATCTACGGNATCCTCGTAAACCAAA	19

Search completed: August 30, 2005, 16:45:49  
Job time : 861 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 30, 2005, 12:51:04 ; Search time 4252 Seconds  
(without alignments)  
9945.767 Million cell updates/sec

Title: US-09-979-549-2

Perfect score: 1111

Sequence: 1 gaagtcagaagccgttcag.....ccgagaagagagagagag 1111

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST.\*

- 1: gb\_est1.\*
- 2: gb\_est2.\*
- 3: gb\_hic.\*
- 4: gb\_est3.\*
- 5: gb\_est4.\*
- 6: gb\_est5.\*
- 7: gb\_est6.\*
- 8: gb\_gss1.\*
- 9: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	105.6	9.5	1061	7	CK213031	CK213031 FGAS02493
2	99	8.9	509	6	CD867573	CD867573 AZO2.106K
3	99	8.9	570	6	CD867174	CD867174 AZO2.105J
4	92	8.3	594	6	CA593541	CA593541 wpaic.pk0
5	92	8.3	631	6	CD938039	CD938039 OV.108O07
6	90	8.1	624	6	CD865240	CD865240 AZO2.0731
7	88	7.9	430	6	CA712930	CA712930 wdk3c.pk0
C 8	79.8	7.2	875	7	CK159678	CK159678 FGAS04115
9	79	7.1	1141	7	CK211149	CK211149 FGAS02298
C 10	78.2	7.0	839	7	CK158306	CK158306 FGAS03954
C 11	77.2	6.9	889	7	CK158740	CK158740 FGAS04003
12	72.4	6.5	628	5	BQ805515	BQ805515 WHE3567.H
C 13	70.6	6.4	867	7	CK158414	CK158414 FGAS03966
14	69.6	6.3	591	6	CD887850	CD887850 G118.106G
15	68.6	6.2	644	6	CD874126	CD874126 AZO3.101H
16	67.6	6.1	630	6	CD930486	CD930486 GR45.111H
17	67.2	6.0	614	6	CA597581	CA597581 wpaic.pk0
18	66.2	6.0	552	6	CA598930	CA598930 WYrlc.pk0
19	66	5.9	643	6	CD884411	CD884411 F1.116120
20	65.2	5.9	567	6	CA717478	CA717478 wdk4c.pk0
21	64.6	5.8	490	6	CD922551	CD922551 G750.103K
22	64.6	5.8	642	7	CF132917	CF132917 WHE4351.G
23	64.4	5.8	712	6	CD894217	CD894217 G118.125L
24	63	5.7	556	6	CA701748	CA701748.wkm2c.pk0

ALIGNMENTS

RESULT 1  
CK213031/c  
LOCUS FGAS024930 Triticum aestivum FGAS: Library 6 CAP GATE 1 Triticum  
DEFINITION aestivum cDNA, mRNA sequence.  
ACCESSION CK213031  
VERSION CK213031.1 GI:39619135  
KEYWORDS EST.  
SOURCE Triticum aestivum (bread wheat)  
ORGANISM Triticum aestivum  
REFERENCE 1 (bases 1 to 1061)  
AUTHORS Allard P., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D., Genswein, B., Graf, R., Gulick, P., Hryciak, L.D., Laroche, A., Links, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilsson, D., Penniket, C., Roach, J.L. and Sarhan, F.  
TITLE Functional Genomics of Abiotic Stress In Wheat and Canola Crops  
JOURNAL Unpublished (2003)  
COMMENT Contact: Wm L Crosby  
Bioinformatics  
University of Saskatchewan, Department of Computer Science  
1C101 Engineering Building, 57 Campus Drive, Saskatoon,  
Saskatchewan, S7N 5A9, Canada  
Tel: 306 966 1769  
Fax: 306 966 2033  
Email: fgas\_escs@usask.ca  
This sequence is the direct result of the Base calling software  
Phred (default parameters). It is the raw base calls. To aid in the  
identification of the high quality insert the software Lucy  
(default parameters) has been run on this sequence. Lucy identified  
the region [27,674].  
Plate: L6B007 row: O column: 02.  
Location/Qualifiers  
1. .1061  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/db\_xref="taxon:4565"  
/clone\_lib="Triticum aestivum FGAS: Library 6 CAP GATE 1"  
/note="Organ: Crown and leaf; Vector: pCMV.SPORT6; Crown  
(50%) and leaf (50%) tissues from wheat cultivar Norstar  
after short exposure times to low temperature in the light  
and in the dark. 12 mRNA populations were combined before  
constructing the library. The first 6 populations: After 7  
days of growth at 20C8 from wheat cultivar Norstar after  
short exposure times to low temperature in the light and

in the dark. 12 mRNA populations were combined before constructing the library. The first 6 populations: After 7 days of growth at 20, wheat plants and green leaf tissue were in the light. 1cm crown sections and green leaf tissue were separately harvested after 1, 3, and 6 hours of low temperature exposure. The last 6 populations: After 7 days of growth at 20C, wheat plants and green leaf tissue were the dark. 1cm crown sections and green leaf tissue were separately harvested after 1, 3, and 6 hours of low temperature exposure. First strand synthesis in this library was done in the presence of methylated dCTP thereby protecting from internal cleavage with NotI. In addition, this library used a primer for second strand synthesis that annealed to an artificial sequence (RNA oligo) added before first strand synthesis. Therefore when sequences from EST generated from this library will be masked for vector and adaptor sequences, an additional masking step will have to be included to mask this RNA oligo that is common to all clones (sequence CGACTGGACGACGAGCAGCTGACATGGAGGAGTAGAAA)."

## ORIGIN

Query Match 9.5%; Score 105.6; DB 7; Length 1061;  
 Best Local Similarity 96.4%; Pred. No. 3.7e-20;  
 Matches 108; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1000 ACAATTTATCTTTGGGAGCGCGCGGATTGGAGACAGAGCCCAAGGCAACAA 1059  
 Db 679 AAAAAATTTATCTTTGGGAGCGCGCGGATTGGAGACAGAGCCCAAGGCAACAA 620

Qy 1060 AGTCGCGGTGAGAATCAACAAGCGGTCTTCCGAGAGAGAGAGAGAG 1111  
 Db 619 AGTCGCGGTGAGAATCAACAAGCGGTCTTCCGAGAGAGAGAGAGAG 568

## RESULT 2

CD867573 509 bp mRNA linear EST 11-JUL-2003  
 LOCUS AZ02.106K05F001108 AZ02 Triticum aestivum cDNA clone AZ02106K05,  
 mRNA sequence.

ACCESSION CD867573  
 VERSION CD867573.1 GI:32551389  
 KEYWORDS EST.

SOURCE Triticum aestivum (bread wheat)  
 ORGANISM Triticum aestivum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poideae; Triticeae; Triticum.

TITLE 1 (bases 1 to 509)  
 JOURNAL Genoplatte.  
 COMMENT Genoplatte, a major partnership french program in plant genomics  
 Unpublished (2003)  
 Contact: Genoplatte  
 Genoplatte  
 93, rue Henri Rochefort 91025 EVRY CEDEX France  
 Tel: 33 1 69 47 54 00  
 Fax: 33 1 69 47 54 10  
 This sequence has been generated in the framework of the french  
 plant genomics programme 'Genoplatte' (<http://www.genoplatte.com>  
 and <http://genoplatte-info.infobiogen.fr>).

## FEATURES

source  
 1..509  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /cultivar="recital"  
 /db\_xref="taxon:4565"  
 /clone="AZ02106K05"  
 /tissue\_type="root"  
 /clone\_lib="AZ02"

## ORIGIN

Query Match 8.9%; Score 99; DB 6; Length 509;  
 Best Local Similarity 100.0%; Pred. No. 3e-18;

Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1013 GGGGAGCGCGCGCGGATTGGAGACAGAGCCCAAGGCAACAAAGTGCCTGAGA 1072  
 Db 1 GGGGAGCGCGCGCGGATTGGAGACAGAGCCCAAGGCAACAAAGTGCCTGAGA 60

Qy 1073 AATCAACAAGCGGTCTTCCGAGAGAGAGAGAGAG 1111  
 Db 61 AATCAACAAGCGGTCTTCCGAGAGAGAGAGAGAG 99

## RESULT 3

CD867174 570 bp mRNA linear EST 11-JUL-2003  
 LOCUS AZ02.105J01F001124 AZ02 Triticum aestivum cDNA clone AZ02105J01,  
 mRNA sequence.

ACCESSION CD867174  
 VERSION CD867174.1 GI:32550990  
 KEYWORDS EST.

SOURCE Triticum aestivum (bread wheat)  
 ORGANISM Triticum aestivum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poideae; Triticeae; Triticum.

TITLE 1 (bases 1 to 570)  
 JOURNAL Genoplatte.  
 COMMENT Genoplatte, a major partnership french program in plant genomics  
 Unpublished (2003)  
 Contact: Genoplatte  
 Genoplatte  
 93, rue Henri Rochefort 91025 EVRY CEDEX France  
 Tel: 33 1 69 47 54 00  
 Fax: 33 1 69 47 54 10  
 This sequence has been generated in the framework of the french  
 plant genomics programme 'Genoplatte' (<http://www.genoplatte.com>  
 and <http://genoplatte-info.infobiogen.fr>).

## FEATURES

source  
 1..570  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /cultivar="recital"  
 /db\_xref="taxon:4565"  
 /clone="AZ02105J01"  
 /tissue\_type="root"  
 /clone\_lib="AZ02"

## ORIGIN

Query Match 8.9%; Score 99; DB 6; Length 570;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-18;  
 Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1013 GGGGAGCGCGCGCGGATTGGAGACAGAGCCCAAGGCAACAAAGTGCCTGAGA 1072  
 Db 1 GGGGAGCGCGCGCGGATTGGAGACAGAGCCCAAGGCAACAAAGTGCCTGAGA 60

Qy 1073 AATCAACAAGCGGTCTTCCGAGAGAGAGAGAGAG 1111  
 Db 61 AATCAACAAGCGGTCTTCCGAGAGAGAGAGAGAG 99

## RESULT 4

CA593541 594 bp mRNA linear EST 21-NOV-2002  
 LOCUS wpalc.pk002.p19 wpalc Triticum aestivum cDNA clone wpalc.pk002.p19  
 DEFINITION 5', end, mRNA sequence.

ACCESSION CA593541  
 VERSION CA593541.1 GI:25143327  
 KEYWORDS EST.

SOURCE Triticum aestivum (bread wheat)  
 ORGANISM Triticum aestivum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poideae; Triticeae; Triticum.



```
REFERENCE
AUTHORS   1 (bases 1 to 594)
           Tingey,S.V., Moore,G., Griffiths,S., Powell,W., Wolters,P.,
           Dolan,M., Hainey,C., Miao,G., Caraher,N. and Hanafey,M.K.
TITLE      DuPont Wheat cDNA Sequence in collaboration with the John Innes
           Center 1
JOURNAL    Unpublished (2002)
COMMENT    Contact: Scott V. Tingey
           Crop Genetics
           E. I. DuPont de Nemours and Company
           1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
           Tel: 302-631-2602
           Fax: 302-631-2607
           Email: Scott.V.Tingey@USA.dupont.com
           Seq primer: M13.
           Location/Qualifiers
FEATURES   1..594
           source
           1..594
           /organism="Triticum aestivum"
           /mol_type="mRNA"
           /db_xref="taxon:4565"
           /clone="wpalc.pk002.p19"
           /tissue_type="anthers"
           /lab_host="DH10B"
           /clone_lib="wpalc"
           /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
           XhoI; Wheat (Triticum aestivum) pre-meiotic anthers JIC"
ORIGIN
Query Match      8.3%; Score 92; DB 6; Length 594;
Best Local Similarity 100.0%; Pred. No. 4.2e-16;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1020 CGGCGCGGATTGGAGACAGAGCCCAAGGCAACAAAGTGCCTGAGAAATCAAC 1079
Db 1 CGGCGCGGATTGGAGACAGAGCCCAAGGCAACAAAGTGCCTGAGAAATCAAC 60

Qy 1080 AAGCGGTGCTTCCGAGAGAGAGAGAGAG 1111
Db 61 AAGCGGTGCTTCCGAGAGAGAGAGAGAGAG 92

RESULT 5
LOCUS      CD938039                631 bp      mRNA      linear      EST 15-JUL-2003
DEFINITION OV.108007F010206 OV Triticum aestivum cDNA clone OV108007, mRNA
sequence.
ACCESSION  CD938039
VERSION     CD938039.1 GI:32785547
KEYWORDS   Triticum aestivum (bread wheat)
SOURCE      Triticum aestivum
ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Triticum.
REFERENCE   1 (bases 1 to 631)
AUTHORS     Genoplante.
TITLE       Genoplante, a major partnership french program in plant genomics
JOURNAL     Unpublished (2003)
COMMENT     Contact: Genoplante
           Genoplante
           93, rue Henri Rochefort 91025 EVRY CEDEX France
           Tel: 33 1 69 47 54 00
           Fax: 33 1 69 47 54 10
           This sequence has been generated in the framework of the french
           plant genomics programme 'Genoplante' (http://www.genoplante.com
           and http://genoplante-info.infobiogen.fr).
           Location/Qualifiers
FEATURES     1..631
           source
           1..631
           /organism="Triticum aestivum"
           /mol_type="mRNA"
           /cultivar="recital"
           /db_xref="taxon:4565"
           /clone="OV108007"
           /tissue_type="ovary"

REFERENCE   1 (bases 1 to 594)
AUTHORS     Tingey,S.V., Moore,G., Griffiths,S., Powell,W., Wolters,P.,
           Dolan,M., Hainey,C., Miao,G., Caraher,N. and Hanafey,M.K.
TITLE      DuPont Wheat cDNA Sequence in collaboration with the John Innes
           Center 1
JOURNAL    Unpublished (2002)
COMMENT    Contact: Scott V. Tingey
           Crop Genetics
           E. I. DuPont de Nemours and Company
           1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
           Tel: 302-631-2602
           Fax: 302-631-2607
           Email: Scott.V.Tingey@USA.dupont.com
           Seq primer: M13.
           Location/Qualifiers
FEATURES   1..594
           source
           1..594
           /organism="Triticum aestivum"
           /mol_type="mRNA"
           /db_xref="taxon:4565"
           /clone="wpalc.pk002.p19"
           /tissue_type="anthers"
           /lab_host="DH10B"
           /clone_lib="wpalc"
           /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
           XhoI; Wheat (Triticum aestivum) pre-meiotic anthers JIC"
ORIGIN
Query Match      8.3%; Score 92; DB 6; Length 594;
Best Local Similarity 100.0%; Pred. No. 4.2e-16;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1020 CGGCGCGGATTGGAGACAGAGCCCAAGGCAACAAAGTGCCTGAGAAATCAAC 1079
Db 1 CGGCGCGGATTGGAGACAGAGCCCAAGGCAACAAAGTGCCTGAGAAATCAAC 60

Qy 1080 AAGCGGTGCTTCCGAGAGAGAGAGAGAG 1111
Db 61 AAGCGGTGCTTCCGAGAGAGAGAGAGAGAG 92

RESULT 6
LOCUS      CD865240                624 bp      mRNA      linear      EST 11-JUL-2003
DEFINITION AZO2.073123F000912 AZO2 Triticum aestivum cDNA clone AZO2073123,
mRNA sequence.
ACCESSION  CD865240
VERSION     CD865240.1 GI:32549056
KEYWORDS   Triticum aestivum (bread wheat)
SOURCE      Triticum aestivum
ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Triticum.
REFERENCE   1 (bases 1 to 624)
AUTHORS     Genoplante.
TITLE       Genoplante, a major partnership french program in plant genomics
JOURNAL     Unpublished (2003)
COMMENT     Contact: Genoplante
           Genoplante
           93, rue Henri Rochefort 91025 EVRY CEDEX France
           Tel: 33 1 69 47 54 00
           Fax: 33 1 69 47 54 10
           This sequence has been generated in the framework of the french
           plant genomics programme 'Genoplante' (http://www.genoplante.com
           and http://genoplante-info.infobiogen.fr).
           Location/Qualifiers
FEATURES     1..624
           source
           1..624
           /organism="Triticum aestivum"
           /mol_type="mRNA"
           /cultivar="recital"
           /db_xref="taxon:4565"
           /clone="AZO2073123"
           /tissue_type="root"
           /clone_lib="AZO2"

ORIGIN
Query Match      8.1%; Score 90; DB 6; Length 624;
Best Local Similarity 100.0%; Pred. No. 1.7e-15;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1022 GCGCGGATTGGAGACAGAGCCCAAGGCAACAAAGTGCCTGAGAAATCAACAA 1081
Db 1 GCGCGGATTGGAGACAGAGCCCAAGGCAACAAAGTGCCTGAGAAATCAACAA 60

Qy 1082 GCGGTGCTTCCGAGAGAGAGAGAGAGAG 1111
Db 61 GCGGTGCTTCCGAGAGAGAGAGAGAGAGAG 90

RESULT 7
LOCUS      CA712930                430 bp      mRNA      linear      EST 26-NOV-2002
DEFINITION wdk3c.pk008.e19 wdk3c Triticum aestivum cDNA clone wdk3c.pk008.e19
5' end, mRNA sequence.
ACCESSION  CA712930
VERSION     CA712930.1 GI:25434723
KEYWORDS   EST.
```

```
ORIGIN
Query Match      8.3%; Score 92; DB 6; Length 631;
Best Local Similarity 100.0%; Pred. No. 4.3e-16;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1020 CGGCGCGGATTGGAGACAGAGCCCAAGGCAACAAAGTGCCTGAGAAATCAAC 1079
Db 1 CGGCGCGGATTGGAGACAGAGCCCAAGGCAACAAAGTGCCTGAGAAATCAAC 60

Qy 1080 AAGCGGTGCTTCCGAGAGAGAGAGAGAG 1111
Db 61 AAGCGGTGCTTCCGAGAGAGAGAGAGAGAG 92

RESULT 6
LOCUS      CD865240                624 bp      mRNA      linear      EST 11-JUL-2003
DEFINITION AZO2.073123F000912 AZO2 Triticum aestivum cDNA clone AZO2073123,
mRNA sequence.
ACCESSION  CD865240
VERSION     CD865240.1 GI:32549056
KEYWORDS   Triticum aestivum (bread wheat)
SOURCE      Triticum aestivum
ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Triticum.
REFERENCE   1 (bases 1 to 624)
AUTHORS     Genoplante.
TITLE       Genoplante, a major partnership french program in plant genomics
JOURNAL     Unpublished (2003)
COMMENT     Contact: Genoplante
           Genoplante
           93, rue Henri Rochefort 91025 EVRY CEDEX France
           Tel: 33 1 69 47 54 00
           Fax: 33 1 69 47 54 10
           This sequence has been generated in the framework of the french
           plant genomics programme 'Genoplante' (http://www.genoplante.com
           and http://genoplante-info.infobiogen.fr).
           Location/Qualifiers
FEATURES     1..624
           source
           1..624
           /organism="Triticum aestivum"
           /mol_type="mRNA"
           /cultivar="recital"
           /db_xref="taxon:4565"
           /clone="AZO2073123"
           /tissue_type="root"
           /clone_lib="AZO2"

ORIGIN
Query Match      8.1%; Score 90; DB 6; Length 624;
Best Local Similarity 100.0%; Pred. No. 1.7e-15;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1022 GCGCGGATTGGAGACAGAGCCCAAGGCAACAAAGTGCCTGAGAAATCAACAA 1081
Db 1 GCGCGGATTGGAGACAGAGCCCAAGGCAACAAAGTGCCTGAGAAATCAACAA 60

Qy 1082 GCGGTGCTTCCGAGAGAGAGAGAGAGAG 1111
Db 61 GCGGTGCTTCCGAGAGAGAGAGAGAGAGAG 90

RESULT 7
LOCUS      CA712930                430 bp      mRNA      linear      EST 26-NOV-2002
DEFINITION wdk3c.pk008.e19 wdk3c Triticum aestivum cDNA clone wdk3c.pk008.e19
5' end, mRNA sequence.
ACCESSION  CA712930
VERSION     CA712930.1 GI:25434723
KEYWORDS   EST.
```

```

SOURCE      Triticum aestivum (bread wheat)
ORGANISM    Triticum aestivum
REFERENCE   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS     Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
            1 (bases 1 to 430)
            Tingey, S.V., Powell, W., Walters, P., Dolan, M., Hainey, C., Yuan, Z.,
            Miao, G., Caraher, N. and Hanafey, M.K.
            DuPont Wheat cDNA Sequence
            Unpublished (2002)
            Contact: Scott V. Tingey
            Crop Genetics
            E. I. DuPont de Nemours and Company
            1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
            Tel: 302-631-2602
            Fax: 302-631-2607
            Email: Scott.V.Tingey@USA.dupont.com
            Seq primer: M13.
FEATURES    Location/Qualifiers
            1..430
            /organism="Triticum aestivum"
            /mol_type="mRNA"
            /db_xref="taxon:4565"
            /clone="wdk3c.pk008.e19"
            /tissue_type="kernel"
            /lab_host="DH10B"
            /clone_lib="wdk3c"
            /note="Vector: pBluescript SK+; Site 1: EcoRI; Site 2:
            XhoI; Wheat (Triticum aestivum L.) developing kernel, 14
            days after anthesis."
ORIGIN
Query Match      7.9%; Score 88; DB 6; Length 430;
Best Local Similarity 100.0%; Pred. No. 6.3e-15;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1024 CCGCGATTGGAGACAGAGCCCAAGGCAACAAAGTGGCGTGAGAAATCAACAAGC 1083
      |||
      1 CCGGATTGGAGACAGAGCCCAAGGCAACAAAGTGGCGTGAGAAATCAACAAGC 60
      |||
      61 GGTGCTTCCGAGAGAGAGAGAGAGAG 88

Qy 1084 GGTGCTTCCGAGAGAGAGAGAGAGAG 1111
      |||
      61 GGTGCTTCCGAGAGAGAGAGAGAGAG 88

Db 61 GGTGCTTCCGAGAGAGAGAGAGAGAG 88

RESULT 8
CK159678/c
LOCUS      CK159678
DEFINITION CK159678 875 bp mRNA linear EST 05-DEC-2003
            FGAS041156 Triticum aestivum FGAS: TaLts Triticum aestivum cDNA,
            mRNA sequence.
ACCESSION  CK159678
VERSION     CK159678.1 GI:38986084
KEYWORDS   EST.
SOURCE     Triticum aestivum (bread wheat)
ORGANISM   Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
            1 (bases 1 to 875)
            Allard, F., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,
            Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Larocche, A.,
            Links, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilsson, D.,
            Penniket, C., Roach, J.L. and Sarhan, F.
            Functional Genomics of Abiotic Stress In Wheat and Canola Crops
            Unpublished (2003)
            Contact: Wm L Crosby
            Bioinformatics
            University of Saskatchewan, Department of Computer Science
            1C101 Engineering Building, 57 Campus Drive, Saskatoon,
            Saskatchewan, S7N 5A9, Canada
            Tel: 306 966 1769
            Fax: 306 966 2033
            Email: fgas_est@cs.usask.ca

TITLE      Functional Genomics of Abiotic Stress In Wheat and Canola Crops
JOURNAL
COMMENT    This sequence is the direct result of the Base calling software
            Phred (default parameters). It is the raw base calls. To aid in the
            identification of the high quality insert the software Lucy
            (default parameters) has been run on this sequence. Lucy identified
            the region [15,726].
            Plate: 16B001 row: D column: 09.
FEATURES    Location/Qualifiers
            1..1141
            /organism="Triticum aestivum"

```

```

This sequence is the direct result of the Base calling software
Phred (default parameters). It is the raw base calls. To aid in the
identification of the high quality insert the software Lucy
(default parameters) has been run on this sequence. Lucy identified
the region [123,778].
Plate: TaLts540 row: B column: 16.
FEATURES    Location/Qualifiers
            1..875
            /organism="Triticum aestivum"
            /mol_type="mRNA"
            /cultivar="wheat line PI 178383"
            /db_xref="taxon:4565"
            /lab_host="DH5 alpha"
            /clone_lib="Triticum aestivum FGAS: TaLts5"
            /note="Organ: Crown; Vector: pGEM-T; SSH (suppression
            subtractive hybridization) cDNA library from genotype
            PI178383 cold hardened at 2 C for 21 days and 49 days
            (equal amount of cDNA pooled together before subtraction,
            tester) and subtracted against genotype Norstar cold
            hardened at 2 C for 1 day (24 H) (driver). Modified Smart
            cDNA (Clontech) priming and non-directional cloning"
ORIGIN
Query Match      7.2%; Score 79.8; DB 7; Length 875;
Best Local Similarity 91.3%; Pred. No. 2.3e-12;
Matches 84; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1020 CGSGCGGATTGGAGACAGAGCCCAAGGCAACAAAGTGGCGTGAGAAATCAAC 1079
      |||
      766 CGSGCGGATTGGATCAACACAGAGCCCAAGGCAACAAAGTGGCGTGAGAAATCAAC 707

Qy 1080 AAGCGGTGCTTCCGAGAGAGAGAGAGAGAG 1111
      |||
      706 AAGCGGTGCTTCCGAGAGAGAGAGAGAGAG 675

Db 706 AAGCGGTGCTTCCGAGAGAGAGAGAGAGAG 675

RESULT 9
CK211149/c
LOCUS      CK211149
DEFINITION CK211149 1141 bp mRNA linear EST 09-DEC-2003
            FGAS022985 Triticum aestivum FGAS: Library 6 CAP GATE 1 Triticum
            aestivum cDNA, mRNA sequence.
ACCESSION  CK211149
VERSION     CK211149.1 GI:39617258
KEYWORDS   EST.
SOURCE     Triticum aestivum (bread wheat)
ORGANISM   Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
            1 (bases 1 to 1141)
            Allard, F., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,
            Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Larocche, A.,
            Links, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilsson, D.,
            Penniket, C., Roach, J.L. and Sarhan, F.
            Functional Genomics of Abiotic Stress In Wheat and Canola Crops
            Unpublished (2003)
            Contact: Wm L Crosby
            Bioinformatics
            University of Saskatchewan, Department of Computer Science
            1C101 Engineering Building, 57 Campus Drive, Saskatoon,
            Saskatchewan, S7N 5A9, Canada
            Tel: 306 966 1769
            Fax: 306 966 2033
            Email: fgas_est@cs.usask.ca
            This sequence is the direct result of the Base calling software
            Phred (default parameters). It is the raw base calls. To aid in the
            identification of the high quality insert the software Lucy
            (default parameters) has been run on this sequence. Lucy identified
            the region [15,726].
            Plate: 16B001 row: D column: 09.
FEATURES    Location/Qualifiers
            1..1141
            /organism="Triticum aestivum"

```

/mol\_type="mRNA"  
 /db\_xref="taxon:4565"  
 /clone\_lib="Triticum aestivum FGAS: Library 6 CAP GATE 1"  
 /note="Organ: Crown and leaf; Vector: pCMV.SPORT6; Crown (50%) and leaf (50%) tissues from wheat cultivar Norstar after short exposure times to low temperature in the light and in the dark. 12 mRNA populations were combined before constructing the library. The first 6 populations: After 7 days of growth at 20C from wheat cultivar Norstar after short exposure times to low temperature in the light and in the dark. 12 mRNA populations were combined before constructing the library. The first 6 populations: After 7 days of growth at 20, wheat plants were transferred to 4C in the light. 1cm crown sections and green leaf tissue were separately harvested after 1, 3, and 6 hours of low temperature exposure. The last 6 populations: After 7 days of growth at 20C, wheat plants were transferred to 4C in the dark. 1cm crown sections and green leaf tissue were separately harvested after 1, 3, and 6 hours of low temperature exposure. First strand synthesis in this library was done in the presence of methylated dCTP thereby protecting from internal cleavage with NotI. In addition, this library used a primer for second strand synthesis that annealed to an artificial sequence (RNA oligo) added before first strand synthesis. Therefore when sequences from EST generated from this library will be masked for vector and adaptor sequences, an additional masking step will have to be included to mask this RNA oligo that is common to all clones (sequence CGACTGGAGCAGCAGGACACTGACATGGCTAGAGGAGTAGAAA)."

## ORIGIN

Query Match 7.1%; Score 79; DB 7; Length 1141;  
 Best Local Similarity 91.4%; Pred. No. 4.4e-12;  
 Matches 96; Conservative 0; Mismatches 5; Indels 4; Gaps 1;

QY 1007 ATCTTGGGAGGCGCGGATTTGGAGACAGAGCCCAAGGCAACAAAGTGCGC 1066  
 |||  
 Db 671 ATATTTGGGAGGCGCGGATTTGGAGACAGAGCCCAAGGCAACAAAGTGCGC 612  
 |||

QY 1067 GTGAGAAATCAACAAGCGGTGCTTCCGAGAGAGAGAGAGAG 1111  
 |||  
 Db 611 GTGAGAAAT---AAGCGGTGCTTCCGAGAGAGAGAGAGAG 571  
 |||

RESULT 10  
 CK158306/c  
 LOCUS  
 DEFINITION FGAS039543 Triticum aestivum FGAS: Talt5 Triticum aestivum cDNA,  
 mRNA sequence.  
 CK158306  
 CK158306.1 GI:38983324  
 EST.  
 Triticum aestivum (bread wheat)  
 ORGANISM

REFERENCE  
 AUTHORS  
 Triticum aestivum (bread wheat)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poideae; Triticeae; Triticum.

1 (bases 1 to 839)  
 Allard, F., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,  
 Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Laroche, A.,  
 Links, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilson, D.,  
 Penniket, C., Roach, J.L. and Sarhan, F.  
 Functional Genomics of Abiotic Stress In Wheat and Canola Crops  
 Unpublished (2003)  
 Contact: Wm L Crosby  
 Bioinformatics  
 University of Saskatchewan, Department of Computer Science  
 1C101 Engineering Building, 57 Campus Drive, Saskatoon,  
 Saskatchewan, S7N 5A9, Canada  
 Tel: 306 966 1769  
 Fax: 306 966 2033  
 Email: fgas\_estcs.usask.ca

TITLE  
 JOURNAL  
 COMMENT

This sequence is the direct result of the Base calling software Phred (default parameters). It is the raw base calls. To aid in the identification of the high quality insert the software Lucy (default parameters) has been run on this sequence. Lucy identified the region [127,643].  
 Plate: Talt534 row: A column: 15.

FEATURES  
 source

Location/Qualifiers  
 1..839  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /cultivar="Wheat line PI 178383"  
 /db\_xref="taxon:4565"  
 /lab\_host="DH5 alpha"  
 /clone\_lib="Triticum aestivum FGAS: Talt5"  
 /note="Organ: Crown; Vector: pGEM-T; SSH (suppression subtractive hybridization) cDNA library from genotype PI178383 cold hardened at 2 C for 21 days and 49 days (equal amount of cDNA pooled together before subtraction, tester) and subtracted against genotype Norstar cold hardened at 2 C for 1 day (24 H) (driver). Modified Smart cDNA (Clontech) priming and non-directional cloning"

## ORIGIN

Query Match 7.0%; Score 78.2; DB 7; Length 839;  
 Best Local Similarity 90.2%; Pred. No. 7e-12;  
 Matches 83; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1020 CGGCGCGGATTTGGAGACAGAGCCCAAGGCAACAAAGTGCGCGTGAATAATCAAC 1079  
 |||  
 Db 770 CGGCGGATTTATACCGAGAGCCCAAGGCAACAAAGTGCGCGTGAATAATCAAC 711  
 |||

QY 1080 AAGCGGTGCTTCCGAGAGAGAGAGAGAG 1111  
 |||  
 Db 710 AAGCGGTGCTTCCGAGAGAGAGAGAGAGAG 679  
 |||

RESULT 11  
 CK158740/c  
 LOCUS

DEFINITION FGAS040031 Triticum aestivum FGAS: Talt5 Triticum aestivum cDNA,  
 mRNA sequence.  
 CK158740  
 CK158740.1 GI:38984191  
 EST.

KEYWORDS  
 SOURCE  
 ORGANISM  
 Triticum aestivum (bread wheat)

REFERENCE  
 AUTHORS  
 Triticum aestivum (bread wheat)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poideae; Triticeae; Triticum.

1 (bases 1 to 889)  
 Allard, F., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,  
 Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Laroche, A.,  
 Links, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilson, D.,  
 Penniket, C., Roach, J.L. and Sarhan, F.  
 Functional Genomics of Abiotic Stress In Wheat and Canola Crops  
 Unpublished (2003)  
 Contact: Wm L Crosby  
 Bioinformatics  
 University of Saskatchewan, Department of Computer Science  
 1C101 Engineering Building, 57 Campus Drive, Saskatoon,  
 Saskatchewan, S7N 5A9, Canada  
 Tel: 306 966 1769  
 Fax: 306 966 2033  
 Email: fgas\_estcs.usask.ca

TITLE  
 JOURNAL  
 COMMENT

This sequence is the direct result of the Base calling software Phred (default parameters). It is the raw base calls. To aid in the identification of the high quality insert the software Lucy (default parameters) has been run on this sequence. Lucy identified the region [125,709].  
 Plate: Talt535 row: O column: 22.

FEATURES  
 source

Location/Qualifiers  
 1..889  
 /organism="Triticum aestivum"

```

/mol_type="mRNA"
/cultivar="Wheat line PI 178383"
/db_xref="taxon:4565"
/lab_host="DH5 alpha"
/clone_lib="Triticum aestivum FGAS: TaLT5"
/notes="Organ: Crown; Vector: pGEM-T; SSH (suppression subtractive hybridization) cDNA library from genotype PI178383 cold hardened at 2 C for 21 days and 49 days (equal amount) of cDNA pooled together before subtraction, tester) and subtracted against genotype Norstar cold hardened at 2 C for 1 day (24 H)(driver). Modified Smart cDNA (Clontech)priming and non-directional cloning"

ORIGIN
Query Match 6.9%; Score 77.2; DB 7; Length 889;
Best Local Similarity 90.1%; Pred. No. 1.4e-11;
Matches 82; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1021 GGGCGGGATTGGAGACAGAGCCGACAAAGGCAACAAAGTGGCGGTGAGAAATCAACA 1080
Db 844 GGGCAGTGTATCAACGACAGAGCCGACAAAGGCAACAAAGTGGCGGTGAGAAATCAACA 785

Qy 1081 AGCGTCTTCCGAGAGAGAGAGAGAG 1111
Db 784 AGCGTCTTCCGAGAGAGAGAGAGAGAG 754

RESULT 12
BQ05515 628 bp mRNA linear EST 31-JUL-2002
LOCUS WHE3567_H02_0032S Wheat developing grains cDNA library Triticum
DEFINITION aestivum cDNA clone WHE3567_H02_003, mRNA sequence.
ACCESSION BQ05515
VERSION BQ05515.1 GI:22029724
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Triticum.
1 (bases 1 to 628)
Altenbach, S., Anderson, O.D., Chao, S., Chin, A., Close, T.J.,
Cronin, K., Crossman, C., Fenton, R.D., Lazo, G.R., Pham, J.,
Rausch, C.J., Wilson, C. and Woo, J.
The structure and function of the expressed portion of the wheat
genomes - Developing grains cDNA library
Unpublished (2002)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 510595773
Fax: 5105595818
Email: oanderson@w.usda.gov
Sequences have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: SK primer.

FEATURES
Location/Qualifiers
1..628
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Butte 86"
/db_xref="taxon:4565"
/clone="WHE3567_H02_003"
/tissue_type="whole grains"
/dev_stage="3-44 days post anthesis seed"
/lab_host="E. coli SOLR"
/clone_lib="Wheat developing grains cDNA library"
/notes="Vector: Lambda ZAP II, excised phagemid; Site 1:
EcoRI; Plants were grown under six following different
environmental regimes in greenhouse, Environment 1)
24oC/17oC day/night, well-watered, with post-anthesis
fertilizer, Environment 2) 24oC/17oC day/night,
well-watered, without post-anthesis fertilizer,
Environment 3) 37oC/17oC day/night, well-watered, with
post-anthesis fertilizer, Environment 4) 37oC/17oC
day/night, well-watered, without post-anthesis fertilizer,
Environment 5) 37oC/17oC day/night plus drought, with
post-anthesis fertilizer, Environment 6) 37oC/17oC
day/night plus drought, without post-anthesis fertilizer,
developing wheat grains from the following were excised
and frozen in liquid nitrogen, Environment 1 at 3, 5, 7,
8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44 DPA Environment
2 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44
DPA Environment 3 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28,
32, 34 DPA Environment 4 at 3, 5, 7, 8, 10, 12, 16, 20,
24, 28, 32, 34 DPA Environment 5 at 3, 5, 7, 8, 10, 12,
16, 20, 24, 28, 30 DPA Environment 6 at 3, 5, 7, 8, 10,
12, 16, 20, 24, 28, 30 DPA and total RNA was prepared by
S. Altenbach and K. Cronin at USDA-ARS, Albany, CA. A
cDNA library was made using poly (A) RNA, and the cDNA
clones in vivo excised to give pBluescript SK(-)
phagemids in the TJ Close lab (Chin, Close, Fenton) at
the University of California, Riverside. Plasmid DNA
preparations and DNA sequencing were performed in the OD
Anderson lab (others)."
```

```
source
1. .867
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Wheat line PI 178383"
/db_xref="taxon:4565"
/lab_host="DMS alpha"
/clone_lib="Triticum aestivum FGAS: Talt5"
/notes="Organ: Crown; Vector: pGEM-T; SSH (suppression subtractive hybridization) cDNA library from genotype PI178383 cold hardened at 2 C for 21 days and 49 days (equal amount of cDNA pooled together before subraction, tester) and subracted against genotype Norstar cold hardened at 2 C for 1 day (24 H)(driver). Modified Smart cDNA (Clontech)priming and non-directional cloning"

ORIGIN
Query Match 6.4%; Score 70.6; DB 7; Length 867;
Best Local Similarity 89.4%; Pred. No. 1.4e-09;
Matches 76; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1027 GGATTGGAGACAGAGCCCAACAAAGTGGCGGTGAGAAATCAACAAGCGGT 1086
Db 837 GGTATCAACGACAGAGCCCAACAAAGCGGTGAGAAATCAACAAGCCGT 778

Qy 1087 GCTTCCGAGAGAGAGAGAGAG 1111
Db 777 GCTTCCGAGAGAGAGAGAGAG 753

RESULT 14
CD887850
LOCUS G118.106G24F010606 G118 Triticum aestivum cDNA clone G118106G24,
DEFINITION mRNA sequence.
ACCESSION CD887850
VERSION 1 GI:32654563
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Triticum.
1 (bases 1 to 591)
Genoplante, a major partnership french program in plant genomics
Unpublished (2003)
Contact: Genoplante
Genoplante
93, rue Henri Rochefort 91025 EVRY CEDEX France
Tel: 33 1 69 47 54 00
Fax: 33 1 69 47 54 10
This sequence has been generated in the framework of the french
plant genomics programme 'Genoplante' (http://www.genoplante.com
and http://genoplante-info.infobiogen.fr).

FEATURES
source
1. .591
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="recital"
/db_xref="taxon:4565"
/clone="G118106G24"
/tissue_type="grain (118 degrees per day after
pollination)"
/clone_lib="G118"

ORIGIN
Query Match 6.3%; Score 69.6; DB 6; Length 591;
Best Local Similarity 91.5%; Pred. No. 2.6e-09;
Matches 86; Conservative 0; Mismatches 4; Indels 4; Gaps 1;

Qy 1018 GCGGGCCGGATTGGAGACAGAGCCCAACAAAGTGGCGGTGAGAAATCA 1077
Db 11 GCGAGCCGGATTGGAGACAGAGCCCAACAAAGTGGCGGTGAGAAAT-- 68

source
1. .591
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="recital"
/db_xref="taxon:4565"
/clone="G118106G24"
/tissue_type="grain (118 degrees per day after
pollination)"
/clone_lib="G118"

ORIGIN
Query Match 6.2%; Score 68.6; DB 6; Length 644;
Best Local Similarity 91.4%; Pred. No. 5.3e-09;
Matches 85; Conservative 0; Mismatches 4; Indels 4; Gaps 1;

Qy 1019 GCGGGCCGGATTGGAGACAGAGCCCAACAAAGTGGCGGTGAGAAATCA 1078
Db 11 GCGAGCCGGATTGGAGACAGAGCCCAACAAAGTGGCGGTGAGAAAT-- 67

Qy 1079 CAAGCGGTGCTTGGCCGAGAGAGAGAGAGAG 1111
Db 68 -AAGCGGTGCTTGGCCGAGAGAGAGAGAGAG 99

Search completed: August 30, 2005, 16:27:31
Job time : 4256 secs
```

